

MARCIN DUDZIŃSKI (Warszawa)
KONRAD FURMAŃCZYK (Warszawa)

Procedury jednoczesnego testowania wielu hipotez i ich zastosowania w analizie mikromacierzy DNA

Streszczenie. W naszej pracy rozważamy różne podejścia do problematyki jednoczesnego testowania wielu hipotez zerowych. W tym kontekście omawiamy procedury testowania typu single-step, step-down i step-up. W szczególności, przedstawiamy własności i zastosowania takich miar błędów testowania, jak: $FWER$, k - $FWER$, FDP , FDR , $pFDR$. Wspomniane procedury testowania są intensywnie wykorzystywane w analizie mikromacierzy DNA, która to analiza umożliwiła monitorowanie poziomów ekspresji wielu genów jednocześnie oraz znajduje ostatnio szerokie zastosowania w diagnostyce, leczeniu i badaniach medycznych.

Słowa kluczowe: procedury jednoczesnego testowania wielu hipotez, $FWER$, FDP , FDR , p -wartość, q -wartość, mikromacierze DNA.

1. Wprowadzenie i podstawowe pojęcia. Ostatnie dziesięciolecie to okres niezwykle intensywnych badań statystycznych, zmierzających do poprawy oraz zwiększenia efektywności i oceny jakości procedur jednoczesnego testowania wielu hipotez. Do wzrostu zainteresowania tego rodzaju badaniami przyczynił się w znacznym stopniu szybki rozwój biotechnologii, w szczególności zaś eksperymentów mikromacierzowych dotyczących analizy DNA. Spośród artykułów poświęconych teorii jednoczesnego testowania wielu hipotez warto wymienić prace: Benjaminiego i Hochberga (1995), Storey'a (2002, 2003), Lehmana i Romano (2005), Lehmana, Romano i Shaffer (2005). Z kolei wśród artykułów prezentujących zastosowania wspomnianej tematyki na uwagę zasługują prace: Storey'a i Tibshiraniego (2003), Dudoit, Shaffer i Boldrick (2003) oraz Pollard, Birknera, van der Laana i Dudoit (2005).

W niniejszym artykule przedstawimy najpopularniejsze procedury jednoczesnego testowania wielu hipotez, omówimy również rodzaje błędów, jakie można popełnić przy korzystaniu z tychże procedur. Zaprezentujemy także zastosowania tego typu procedur w analizach mikromacierzowych, przy czym skoncentrujemy się głównie na przedstawieniu tych spośród zastosowań, które można wykorzystać do celów porównań miar ekspresji genów

u osób pochodzących z dwóch różnych, odpowiednio wyselekcjonowanych, grup eksperymentalnych.

W pracy rozważamy problemy jednoczesnego testowania hipotez zerowych H_i , $i = 1, \dots, m$, przy założeniach, że: 1) liczba testowanych hipotez zerowych jest bardzo duża (rzędu setek, tysięcy, dziesiątek tysięcy), 2) dysponujemy odpowiednimi testami weryfikacyjnymi, oddzielnie dla każdej z hipotez zerowych, o statystykach testowych T_i , $i = 1, \dots, m$.

Przez procedurę testową będziemy rozumieć regułę określającą, które spośród hipotez zerowych H_i należałoby przyjąć, a które z nich powinniśmy odrzucić.

Klasycznym podejściem do zagadnień dotyczących jednoczesnego testowania wielu hipotez jest konstrukcja procedury testowej, która zapewnia kontrolę miary, będącej prawdopodobieństwem fałszywego odrzucenia co najmniej jednej hipotezy zerowej. Miarę tę nazwano *familywise error rate* – w skrócie *FWER*.

Niech zatem V oznacza liczbę fałszywych odrzuceń. Wtedy $FWER := \mathbb{P}(V \geq 1)$, zaś procedura kontrolująca $FWER$ przy ustalonym poziomie α to procedura zapewniająca spełnienie warunku $FWER \leq \alpha$.

Nieco przystępniejszym dla Czytelnika może być rozpatrywanie kontroli miary $FWER$ poprzez rozważanie wartości prawdopodobieństw $\mathbb{P}(V < 1)$, tzn. prawdopodobieństw, że V – liczba fałszywie odrzuconych hipotez zerowych – będzie mniejsza od 1, a zatem prawdopodobieństw tego, że żadna spośród weryfikowanych hipotez nie zostanie błędnie odrzucona. Jest bowiem faktem oczywistym, że $FWER = 1 - \mathbb{P}(V < 1)$. Stąd łatwo zauważyć, iż procedura kontrolująca $FWER$ na poziomie α to procedura, która zapewnia spełnienie warunku $\mathbb{P}(V < 1) \geq 1 - \alpha$. Kontrola $FWER$ na poziomie α oznacza tym samym, że prawdopodobieństwo tego, iż żadna hipoteza zerowa nie zostanie fałszywie odrzucona będzie nie mniejsze niż $1 - \alpha$.

Procedura, która kontroluje $FWER$ tylko wtedy, gdy 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.

Podejście do tematyki jednoczesnego testowania skończonej liczby hipotez, polegające na kontroli miary $FWER$, jest jednak obarczone pewnymi wadami. Zasadnicza z nich polega na tym, że moc procedur kontrolujących $FWER$, rozumiana jako zdolność tychże procedur do wykrywania fałszywych hipotez zerowych, maleje wraz ze wzrostem m – liczby weryfikowanych hipotez. Z tej to przyczyny, zamiast procedur kontrolujących $FWER$, postanowiono rozważać procedury kontrolujące miarę błędu testowania o nazwie k - $FWER$, $k \geq 1$, definiowaną jako prawdopodobieństwo błędnego odrzucenia co najmniej k spośród weryfikowanych hipotez zerowych H_i . Stąd

k - $FWER := \mathbb{P}(V \geq k)$, sama zaś kontrola miary k - $FWER$ na poziomie α polega na znalezieniu procedury, przy której relacja k - $FWER \leq \alpha$ będzie spełniona dla wszystkich możliwych konfiguracji prawdziwych i fałszywych hipotez zerowych (dotyczy to mocnej kontroli) lub dla przypadku, gdy wszystkie hipotezy zerowe są prawdziwe (dotyczy to słabej kontroli). Oczywiście jest fakt, że gdy $k = 1$ – wtedy k - $FWER$ to zwykły $FWER$.

Podobnie jak w przypadku kontroli miary $FWER$, również kwestię kontroli miary k - $FWER$ można rozpatrywać, rozważając prawdopodobieństwo $\mathbb{P}(V < k)$, tzn. prawdopodobieństwo, że V – liczba fałszywie odrzuconych hipotez zerowych – będzie mniejsza od k . Mianowicie, jest faktem oczywistym, że k - $FWER = 1 - \mathbb{P}(V < k)$. Stąd w prosty sposób wynika, że procedura kontrolująca k - $FWER$ na poziomie α to taka procedura, która zapewnia spełnienie warunku $\mathbb{P}(V < k) \geq 1 - \alpha$. Kontrola k - $FWER$ na poziomie α oznacza zatem, że prawdopodobieństwo tego, iż mniej niż k hipotez zerowych zostanie fałszywie odrzuconych będzie nie mniejsze od $1 - \alpha$.

Nietrudno zauważyć, że liczba fałszywie odrzucanych hipotez rośnie wraz ze wzrostem liczby wszystkich odrzucanych hipotez. Stąd powstał pomysł, by zamiast kontroli liczby fałszywych odrzuceń badać stosunek liczby fałszywych odrzuceń do liczby wszystkich odrzuceń. W tym celu wprowadzono nową miarę, o nazwie *false discovery proportion* – w skrócie FDP . Jest ona zdefiniowana następująco. Niech: V – liczba fałszywych odrzuceń, R – liczba wszystkich odrzuceń. Wtedy

$$FDP := \begin{cases} V/R, & \text{gdy } R \neq 0, \\ 0, & \text{gdy } R = 0. \end{cases}$$

Kontrola FDP polega na skonstruowaniu procedury, która przy ustalonych liczbach γ, α z przedziału $(0, 1)$ zapewni spełnienie warunku

$$\mathbb{P}(FDP > \gamma) \leq \alpha.$$

Benjamini i Hochberg (1995) zaproponowali zastąpienie procedur kontrolujących FDP procedurami kontrolującymi miarę o angielskiej nazwie *false discovery rate* – w skrócie FDR – określoną jako $FDR := \mathbb{E}(FDP)$, gdzie $\mathbb{E}(\cdot)$ oznacza wartość oczekiwaną. Kontrola FDR na poziomie α polega na znalezieniu procedury, przy której zachodzi relacja

$$FDR \leq \alpha.$$

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

Nietrudno pokazać, iż jeśli wszystkie hipotezy zerowe są prawdziwe, to $FDR = FWER$. Mówimy wtedy, że kontrola FDR jest równoważna kontroli $FWER$ w słabym sensie. W przypadku, gdy nie wszystkie weryfikowane hipotezy są prawdziwe, zachodzi relacja $FDR \leq FWER$.

W tym kontekście warto wspomnieć także o wprowadzonej przez Storey'a (2002) mierze $pFDR$ (*positive false discovery rate*), stanowiącej prostą modyfikację FDR , oraz o związanym z $pFDR$ pojęciu q -wartości. Zarówno o mierze $pFDR$, jak i o terminie q -wartości piszemy szerzej w rozdziale 5.

2. Model statystyczny. Punktem wyjścia naszych rozważań jest model statystyczny, tj. przestrzeń wyposażona w rodzinę $\{\mathbb{P}_\theta, \theta \in \Theta\}$ rozkładów prawdopodobieństwa. Załóżmy, że obserwacje, stanowiące wyniki eksperymentu, będziemy traktować jako wektor losowy $\mathbf{X} = (X_1, \dots, X_n)$, gdzie X_i są zmiennymi (wektorami) losowymi o rozkładzie \mathbb{P}_θ dla pewnego $\theta \in \Theta$, gdzie Θ jest zbiorem parametrów. Każdą z hipotez H_i , $i = 1, \dots, m$, możemy utożsamiać z pewnym podzbiorem $\Theta_i \subset \Theta$ i zapisać w formie następującej $H_i: \theta \in \Theta_i$, $i = 1, \dots, m$.

Niech $M_0 := \{i : H_i \text{ jest prawdziwa}\}$. Wtedy mocna kontrola miary k - $FWER$ na poziomie α oznacza, że

$$\mathbb{P}_\theta(V \geq k) \leq \alpha \text{ dla każdego } \theta \in \bigcap_{i \in M_0} \Theta_i$$

lub równoważnie, że

$$\mathbb{P}_\theta(V < k) \geq 1 - \alpha \text{ dla każdego } \theta \in \bigcap_{i \in M_0} \Theta_i.$$

Z kolei, słaba kontrola miary k - $FWER$ na poziomie α wymaga spełnienia warunku

$$\mathbb{P}_\theta(V \geq k) \leq \alpha \text{ dla każdego } \theta \in \bigcap_{i \in \{1, \dots, m\}} \Theta_i$$

lub równoważnego mu warunku

$$\mathbb{P}_\theta(V < k) \geq 1 - \alpha \text{ dla każdego } \theta \in \bigcap_{i \in \{1, \dots, m\}} \Theta_i.$$

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

Przystępny opis, wraz z intuicją, pojęcia p -wartości podano w podręczniku Koronackiego i Mielniczuka (2004). Przedstawimy teraz pokrótce zaproponowane przez wspomnianych autorów podejście do tego, cieszącego się coraz większą popularnością i coraz powszechniej stosowanego we współczesnej statystyce, pojęcia.

Założmy, że weryfikujemy hipotezę zerową H_0 , wobec hipotezy alternatywnej H'_0 , przy użyciu statystyki testowej T .

Przyjmijmy, że, zamiast weryfikowania hipotezy H_0 przy z góry ustalonym poziomie istotności α (np. $\alpha = 0,05$), chcemy znaleźć minimalny poziom istotności, przy którym zaobserwowana wartość t , statystyki testowej

T , spowoduje podjęcie decyzji o odrzuceniu H_0 . Taki minimalny poziom istotności określamy mianem p -wartości. Zapiśmy to bardziej formalnie. Załóżmy, jak uprzednio, że weryfikujemy hipotezę H_0 , wobec hipotezy H'_0 , wykorzystując zaobserwowaną wartość t , statystyki testowej T , oraz indeksowaną parametrem α rodzinę obszarów krytycznych (zbiorów odrzuceń) S_α , taką, że $\mathbb{P}_{H_0}(T \in S_\alpha) \leq \alpha$, gdzie $\mathbb{P}_{H_0}(\cdot)$ oznacza prawdopodobieństwo wyznaczone przy założeniu prawdziwości hipotezy zerowej H_0 . Wtedy p -wartość definiujemy następująco

$$p(t) := \inf_{\{\alpha: t \in S_\alpha\}} \mathbb{P}_{H_0}(T \in S_\alpha).$$

W celu lepszego zrozumienia wspomnianego terminu podamy p -wartości dla przykładu dotyczącego weryfikacji hipotezy o średniej z rozkładu normalnego o nieznanym odchyleniu standardowym. Wtedy, jak powszechnie wiadomo, stosujemy statystykę testową t -Studenta postaci

$$T = \frac{\bar{X} - \theta_0}{S/\sqrt{n}},$$

gdzie: \bar{X} – średnia z próby, S – odchylenie standardowe z próby, θ_0 – średnia hipotetyczna, n – rozmiar próby.

W przypadku weryfikacji hipotezy zerowej $H_0: \theta = \theta_0$, wobec hipotezy alternatywnej $H'_0: \theta \neq \theta_0$, otrzymujemy, że $S_\alpha = (-\infty, -t_{\alpha, n-1}) \cup (t_{\alpha, n-1}, \infty)$, gdzie $t_{\alpha, n-1}$ spełnia warunek $F(t_{\alpha, n-1}) = 1 - \alpha/2$ (F oznacza tu dystrybucję rozkładu t -Studenta o $n - 1$ stopniach swobody), sama zaś p -wartość dana jest relacją $p(t) = \mathbb{P}_{H_0}(|T| \geq |t|) = 2\mathbb{P}_{H_0}(T \geq |t|)$, przy czym, dla przypomnienia, t oznacza zaobserwowaną wartość statystyki T .

Z kolei, w przypadku weryfikacji hipotezy zerowej $H_0: \theta = \theta_0$, wobec hipotezy alternatywnej $H'_0: \theta > \theta_0$, otrzymujemy, że: $S_\alpha = (t_{2\alpha, n-1}, \infty)$, $p(t) = \mathbb{P}_{H_0}(T \geq t)$.

Wreszcie, gdy weryfikujemy hipotezę zerową $H_0: \theta = \theta_0$, wobec hipotezy alternatywnej $H'_0: \theta < \theta_0$, otrzymujemy, że: $S_\alpha = (-\infty, -t_{2\alpha, n-1})$, $p(t) = \mathbb{P}_{H_0}(T \leq t)$.

Wyznaczenie p -wartości testu jest o tyle wygodne, iż zwalnia nas ono z konieczności arbitralnego doboru poziomu istotności przed przeprowadzeniem testu. Po wyznaczeniu wartości statystyki testowej sprawdzamy, dla jakiego poziomu istotności owa zaobserwowana wartość statystyki testowej jest wartością krytyczną testu, tj. wartością znajdującą się na brzegu obszaru krytycznego. Im mniejsza p -wartość – tym jesteśmy bardziej skłonni do odrzucenia hipotezy zerowej na korzyść hipotezy alternatywnej, innymi słowy – tym silniejsze staje się nasze przeświadczenie o fałszywości hipotezy zerowej i prawdziwości hipotezy alternatywnej.

Częstym błędem popełnianym przy interpretacji p -wartości jest traktowanie jej jako poziomu istotności przeprowadzanego testu. Owa p -wartość

nie może stanowić błędu pierwszego rodzaju, gdyż w istocie swej – jest ona zmienną losową, nie zaś, z góry wybraną bądź ustaloną przed wykonaniem testu, liczbą, jak to ma miejsce w przypadku poziomu istotności testu.

Należy podkreślić, że p -wartość to pojęcie fundamentalne we współczesnej statystyce matematycznej, odgrywające niepoślednią rolę w jej licznych zastosowaniach. Badania statystyczne wykorzystujące znaczenie p -wartości rozwijają się niezwykle intensywnie na przestrzeni ostatnich lat. Ich doniosłość i znaczenie dla obliczeń wykorzystywanych do rozwiązywania problemów stawianych obecnie w naukach biologicznych, a w szczególności w genetyce, jest tak duże, iż nawet miesiące odgrywają tu istotną rolę. Wspomniane badania cechuje coraz większy stopień złożoności i dość skomplikowany, specjalistyczny język. Coraz powszechniej podejmowane są też próby nieklasycznego podejścia do wspomnianych zagadnień. Jako przykład takiego nowatorskiego podejścia posłużyć mogą rozważania z pracy Gelmana (2003), w której zdefiniowano pojęcie p -wartości przy użyciu terminów statystyki bayesowskiej. W nieco innym ujęciu nawiązujemy do tej tematyki również my w niniejszej pracy (zob. podrozdział 5.2).

Warto nadmienić, że błędna interpretacja terminu p -wartości i jego nieprzemyślane stosowanie może prowadzić do wielu błędnych wniosków. O nadużyciach tego rodzaju pisał już jeden z twórców współczesnej teorii wnioskowania statystycznego, Ronald Fisher. Błędy wynikające z nierozważnego interpretowania zagadnienia p -wartości opisano dość szczegółowo w pracy Hubbarda i Bayarri (2003).

W kontekście problematyki jednoczesnego testowania wielu hipotez pojęcie p -wartości zostało bardzo niedawno, w nieco innym od dotychczasowych kształcie, zaprezentowane przez Victor i Hommela (2007).

W naszych rozważaniach będziemy zakładali, że p -wartości p_i , odpowiadające hipotezom zerowym H_i , $i = 1, \dots, m$, spełniają założenie

$$(1) \quad \mathbb{P}_\theta(p_i \leq u) \leq u \text{ dla wszystkich } u \in (0, 1) \text{ i każdego } \theta \in \Theta_i.$$

Rozważmy pojedynczą hipotezę zerową H_i : $\theta \in \Theta_i$. Załóżmy, że dysponujemy rodziną testów hipotezy H_i , indeksowaną przez α , o statystyce testowej T_i i zbiorach odrzuceń S_α , które spełniają warunki:

$$\mathbb{P}_\theta(T_i \in S_\alpha) \leq \alpha \text{ dla wszystkich } 0 < \alpha < 1, \theta \in \Theta_i,$$

oraz

$$S_\alpha \subset S_\beta \text{ dla } \alpha < \beta.$$

Wtedy p -wartość określamy relacją

$$\begin{aligned} p_i &= p_i(T_i) := \inf\{\alpha : T_i \in S_\alpha\} \\ &= \inf\{\alpha : \text{hipoteza } H_i \text{ jest odrzucana na poziomie istotności } \alpha\}. \end{aligned}$$

Warto dodać, że tak zdefiniowana p -wartość p_i nie zależy od parametru θ oraz spełnia warunek (1) (zob. Lehmann i Romano (2005), Lemat 1.1(i)). Podana definicja p -wartości znalazła zastosowanie w wielu parametrycznych i niektórych nieparametrycznych procedurach jednoczesnego testowania wielu hipotez, np. w testach permutacyjnych lub randomizacyjnych (zob. Lehmann i Romano (2005) oraz Romano i Wolf (2005)).

Niech z kolei, μ oznacza dowolną (np. $FWER$, k - $FWER$, FDR) miarę błędu procedury testowania wielu hipotez. Wtedy dopasowaną p -wartość definiujemy następująco

$$\tilde{p}_i := \inf\{\alpha : \text{hipoteza } H_i \text{ jest odrzucana na poziomie błęd } \mu = \alpha\}.$$

3. Procedury kontroli $FWER$ i k - $FWER$. W niniejszym rozdziale omówimy procedury kontroli miar $FWER$ i k - $FWER$ oraz przedstawimy ich najciekawsze własności.

3.1. Procedury typu *single-step*. Procedura typu *single-step* to procedura, wedle której odrzucane są wszystkie te hipotezy zerowe, których p -wartości nie przekraczają pewnej, ustalonej, wartości granicznej, zwanej z angielska *cut-off value*.

Szczególnym przypadkiem procedury typu *single-step* jest procedura *Bonferroniego*. Jest to procedura, która, przy ustalonym poziomie istotności α , 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 procedura *Bonferroniego* zapewnia mocną kontrolę miary $FWER$ na poziomie α , jeśli p -wartości p_i , $i = 1, \dots, m$, spełniają (1).

Warto dodać, iż w przypadku procedury *Bonferroniego*, kontroli $FWER$, dopasowane p -wartości są dane wzorem

$$\tilde{p}_i = \min(mp_i, 1).$$

Procedura *Bonferroniego* ma tę wadę, że jej moc, czyli zdolność do wykrywania hipotez fałszywych, maleje wraz ze wzrostem m – liczby weryfikowanych hipotez. W przypadku bowiem, gdy m jest duże, to α/m – poziom istotności pojedynczego testu – jest mały.

W procedurze *Bonferroniego*, którą wykorzystujemy do kontroli miary $FWER$, porównujemy każdą p -wartość p_i z α/m . Kontrola miary k - $FWER$ umożliwia zwiększenie tej wartości z α/m do $k\alpha/m$. W konsekwencji, procedura kontrolująca k - $FWER$, nazywana często uogólnioną procedurą *Bonferroniego*, jest procedurą większej mocy od zwykłej procedury *Bonferroniego*, kontrolującej $FWER$. Dokładniej, można udowodnić następujące twierdzenie (zob. Lehmann i Romano (2005), Twierdzenie 2.1).

Twierdzenie A. *Załóżmy, że testujemy jednocześnie hipotezy zerowe $H_i : \theta \in \Theta_i$, $i = 1, \dots, m$, a odpowiadające im p -wartości p_i spełniają wa-*

runek (1). Rozważmy procedurę, która odrzuca H_i , jeśli $p_i \leq k\alpha/m$. Wtedy procedura ta zapewnia mocną kontrolę k -FWER na poziomie α .

Dowód. Niech $M_0 := \{i : H_i \text{ jest prawdziwa}\}$, natomiast m_0 oznacza liczbę hipotez prawdziwych wśród m testowanych hipotez. Niech ponadto: $I(A)$ będzie zmienną losową, określoną wzorem

$$I(A) = I(A, \omega) = \begin{cases} 1 & \text{dla } \omega \in A \\ 0 & \text{dla } \omega \notin A, \end{cases}$$

natomiast V oznacza liczbę fałszywych odrzuceń. Wtedy, na mocy nierówności Markowa i założenia (1),

$$\begin{aligned} k\text{-FWER} = \mathbb{P}_\theta(V \geq k) &\leq \frac{\mathbb{E}_\theta(V)}{k} = \frac{\mathbb{E}_\theta\left(\sum_{i \in M_0} I(p_i \leq k\alpha/m)\right)}{k} \\ &= \sum_{i \in M_0} \frac{\mathbb{P}_\theta(p_i \leq k\alpha/m)}{k} \leq \sum_{i \in M_0} \frac{k\alpha/m}{k} = m_0 \frac{\alpha}{m} \leq \alpha. \end{aligned}$$

To kończy dowód Twierdzenia A. □

Inna znana procedura typu single-step to procedura kontroli FWER, o nazwie $\max T$. Jest ona opisana przy użyciu dopasowanych p -wartości postaci

$$\tilde{p}_i = \mathbb{P}\left(\max_{1 \leq j \leq m} |T_j| \geq t_i\right),$$

gdzie prawdopodobieństwo $\mathbb{P}(\cdot)$ jest wyznaczane przy założeniu prawdziwości wszystkich hipotez H_i , natomiast t_i są realizacjami statystyk T_i .

Można pokazać, że procedura $\max T$ zapewnia słabą kontrolę miary FWER (zob. Westfall i Young (1993), s. 42).

3.2. Procedury typu step-down. Niech

$$(2) \quad \alpha_1 \leq \dots \leq \alpha_m$$

będzie niemalejącym ciągiem stałych.

Oznaczmy przez

$$(3) \quad p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)}$$

uporządkowany ciąg p -wartości, zaś przez $H_{(1)}, H_{(2)}, \dots, H_{(m)}$ hipotezy zerowe, którym owe p -wartości odpowiadają.

Procedury typu step-down można opisać w sposób następujący. Jeśli

$$p_{(1)} > \alpha_1,$$

to nie odrzucamy żadnej hipotezy zerowej; w przeciwnym przypadku, gdy

$$(4) \quad p_{(1)} \leq \alpha_1, \dots, p_{(r)} \leq \alpha_r,$$

to odrzucamy hipotezy $H_{(1)}, \dots, H_{(r)}$, gdzie r jest największą liczbą, która spełnia (4).

Realizację procedur step-down rozpoczynamy od odrzucenia hipotezy zerowej, której odpowiada najmniejsza p -wartość, a potem kontynuujemy odrzucanie tych spośród następujących hipotez zerowych, których p -wartości są odpowiednio małe. W zastosowaniach do kontroli miary $FWER$ procedury z rodzaju step-down mają na ogół większą moc od wykorzystywanych wcześniej do tego celu procedur typu single-step.

Jeden z przykładów procedury step-down stanowi procedura *Holma*. Jest to procedura step-down z α_i postaci $\alpha_i = \alpha / (m - i + 1)$. Warto dodać, iż w przypadku procedury *Holma*, kontroli $FWER$, dopasowane p -wartości są dane wzorem (zob. Dudoit, Shaffer i Boldrick (2003), s. 9)

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

Dla celów mocnej kontroli miary k - $FWER$ Lehmann i Romano (2005) udoskonalili wspomnianą wcześniej procedurę *Holma*. Opracowana przez nich procedura, nosząca nazwę uogólnionej procedury *Holma*, to procedura step-down z α_i danymi wzorem

$$(5) \quad \alpha_i = \begin{cases} k\alpha/m, & i \leq k, \\ k\alpha/(m + k - i), & i > k. \end{cases}$$

Zauważmy, że gdy $k = 1$, to uogólniona procedura *Holma*, kontroli miary k - $FWER$, staje się zwykłą procedurą *Holma*, kontroli miary $FWER$.

O zastosowaniu uogólnionej procedury *Holma* do kontroli k - $FWER$ traktuje poniższe twierdzenie (zob. Lehmann i Romano (2005), Twierdzenie 2.2).

Twierdzenie B. *Założmy, że testujemy jednocześnie hipotezy zerowe $H_i: \theta \in \Theta_i$, $i = 1, \dots, m$, a odpowiadające im p -wartości p_i spełniają warunek (1). Wtedy uogólniona procedura *Holma* zapewnia mocną kontrolę k - $FWER$ na poziomie α .*

Dowód. Niech $M_0 := \{i : H_i \text{ jest prawdziwa}\}$, natomiast m_0 oznacza liczbę hipotez prawdziwych wśród m testowanych hipotez. Założmy, że $m_0 \geq k$ (gdyż w przeciwnym przypadku nie ma czego dowodzić). Uporządkujmy p -wartości odpowiadające prawdziwym hipotezom zerowym i oznaczmy je następująco:

$$q_{(1)} \leq \dots \leq q_{(m_0)}.$$

Niech ponadto, j oznacza najmniejszy indeks, taki, że $p_{(j)} = q_{(k)}$. Wtedy

$$(6) \quad k \leq j \leq m - m_0 + k.$$

Komentarza wymaga jedynie druga z powyższych nierówności. Wynika ona z tego, że największa możliwa wartość j jest osiągnięta wówczas, gdy wszystkie najmniejsze p -wartości odpowiadają $m - m_0$ fałszywym hipotezom ze-

rowym, a następane p -wartości odpowiadają m_0 prawdziwym hipotezom zerowym, tzn. gdy $j = m - m_0 + k$.

Ponieważ $p_{(j)} = q_{(k)}$, to uogólniona procedura *Holma* dopuszcza co naj-
mniej k fałszywych odrzuceń wtedy i tylko wtedy, gdy

$$p_{(1)} \leq \alpha_1, p_{(2)} \leq \alpha_2, \dots, p_{(j)} \leq \alpha_j.$$

Zatem, na mocy (5), (6),

$$q_{(k)} = p_{(j)} \leq \alpha_j = \frac{k\alpha}{m + k - j} \leq \frac{k\alpha}{m_0}.$$

To oraz Twierdzenie A implikują, że

$$k\text{-FWER} \leq \mathbb{P}_\theta \left(q_{(k)} \leq \frac{k\alpha}{m_0} \right) \leq \alpha,$$

co kończy dowód Twierdzenia B. □

3.3. Procedury typu step-up. Oznaczmy, tak jak w (2), (3), przez $\{\alpha_i\}$ – niemalejący ciąg stałych oraz przez $\{p_{(i)}\}$ – uporządkowany ciąg p -wartości. Niech z kolei, $H_{(i)}$ oznaczają hipotezy zerowe, którym odpowiadają p -wartości $p_{(i)}$. Procedurę typu step-up można opisać następująco. Jeśli

$$p_{(m)} \leq \alpha_m,$$

to nie przyjmujemy żadnej hipotezy zerowej; w przeciwnym przypadku, gdy

$$(7) \quad p_{(m)} > \alpha_m, \dots, p_{(r+1)} > \alpha_{r+1},$$

to przyjmujemy hipotezy $H_{(r+1)}, \dots, H_{(m)}$, gdzie r jest najmniejszą liczbą, która spełnia (7).

Procedury step-up rozpoczynamy od przyjęcia hipotezy zerowej, której odpowiada największa p -wartość, a następnie kontynuujemy przyjmowanie tych spośród następných hipotez zerowych, których p -wartości są odpowiednio duże.

Dodajmy, iż w pracy Hochberga (1988) podano procedurę step-up, która umożliwia kontrolę miary *FWER* przy założeniu niezależności p -wartości weryfikowanych hipotez.

Warto przy tym nadmienić, iż w przypadku procedury step-up *Hochberga* dopasowane p -wartości są dane wzorem (zob. Dudoit, Shaffer i Boldrick (2003), s. 10)

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

Interesującym rezultatem dotyczącym mocnej kontroli miary k -*FWER*, przy użyciu procedur typu step-up, jest Twierdzenie 3.1 z pracy Romano i Shaikha (2006). Jest ono o tyle ciekawe, że nie zakłada niezależności p -wartości hipotez zerowych H_i .

4. Procedury kontroli FDP i FDR. W tym rozdziale omówimy procedury kontroli miar FDP i FDR oraz zaprezentujemy ich najciekawsze własności.

4.1. Procedura step-down kontroli FDP. Do celów kontroli miary FDP możemy wykorzystać procedurę typu step-down z α_i określonymi wzorem

$$(8) \quad \alpha_i = \frac{([\gamma i] + 1) \alpha}{m + [\gamma i] + 1 - i},$$

gdzie $[x]$ oznacza największą liczbę całkowitą mniejszą od x lub równą x , natomiast γ, α są ustalonymi liczbami z przedziału $(0, 1)$.

Prawdziwe jest następujące twierdzenie (zob. Lehmann i Romano (2005), Twierdzenie 3.1).

Twierdzenie C. Niech: p_1, \dots, p_m – p -wartości odpowiadające hipotezom zerowym $H_i: \theta \in \Theta_i, i = 1, \dots, m$, q_1, \dots, q_{m_0} – p -wartości odpowiadające m_0 hipotezom prawdziwym, r_1, \dots, r_{m-m_0} – p -wartości odpowiadające $m - m_0$ hipotezom fałszywym. Załóżmy, że, gdy $i = 1, \dots, m$,

$$(9) \quad \mathbb{P}_\theta (q_i \leq u | r_1, \dots, r_{m-m_0}) \leq u$$

dla wszystkich $u \in (0, 1)$ i każdego $\theta \in \Theta_i$. Wtedy procedura typu step-down z α_i określonymi tak jak w (8) zapewnia mocną kontrolę FDP, która polega na zachodzeniu warunku

$$\mathbb{P}_\theta (FDP > \gamma) \leq \alpha \text{ dla każdego } \theta \in \bigcap_{i \in M_0} \Theta_i,$$

gdzie $M_0 := \{i : H_i \text{ jest prawdziwa}\}$.

Dowód Twierdzenia C pomijamy.

Jedyny warunek dotyczący zależności p -wartości, zakładany w twierdzeniu C, stanowi warunek (9). Jest on spełniony np. wtedy, gdy p -wartości odpowiadające prawdziwym hipotezom zerowym są niezależne od odpowiednich p -wartości fałszywych hipotez zerowych. W tym przypadku dopuszczalna jest dowolna zależność p -wartości w obrębie każdej z dwóch grup hipotez: prawdziwych i fałszywych. Okazuje się, że warunek (9) można osłabić. Można bowiem, bez czynienia jakichkolwiek, poza warunkiem (1), założeń o zależności p -wartości, skutecznie stosować procedurę step-down, kontroli FDP, zastępując α_i w (8) przez α'_i postaci

$$(10) \quad \alpha'_i = \frac{\alpha_i}{C_{([\gamma s]+1)}},$$

gdzie $C_j = \sum_{i=1}^j 1/i$ (zob. Lehmann i Romano (2005), Twierdzenie 3.3).

Należy jednak stanowczo podkreślić, iż wspomniane osłabienie warunku (9) odbywa się kosztem zmniejszenia mocy procedury.

4.2. Procedura step-up kontroli FDR. W niniejszym podrozdziale przedstawimy procedurę kontroli *FDR*, opracowaną przez Benjaminiego i Hochberga (1995). Przypomnijmy, że *FDR* definiujemy jako wartość oczekiwaną miary *FDP*, tzn.

$$FDR = \mathbb{E}(FDP) = \mathbb{E}\left(\frac{V}{R \vee 1}\right) = \mathbb{E}\left(\frac{V}{R} \mid R > 0\right) \mathbb{P}(R > 0),$$

gdzie $R \vee 1 := \max(R, 1)$.

Zaproponowana przez Benjaminiego i Hochberga procedura kontroli *FDR* jest procedurą step-up z $\alpha_i = i\alpha/m$. W przypadku tej procedury dopasowane *p*-wartości są określone zależnością (zob. Dudoit, Shaffer i Bol-drick (2003), s. 11)

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

O zastosowaniu procedury *Benjaminiego i Hochberga* do kontroli miary *FDR* traktuje poniższe twierdzenie (zob. Benjamini i Hochberg (1995), Twierdzenie 1).

TWIERDZENIE D. *Dla niezależnych statystyk testowych (p-wartości) procedura Benjaminiego i Hochberga zapewnia mocną kontrolę FDR na poziomie istotności α , tzn.*

$$FDR \leq \frac{m_0}{m} \alpha \leq \alpha.$$

Dowód Twierdzenia D pomijamy.

Przewaga miary *FDR* i związanej z jej kontrolą procedury *Benjaminiego i Hochberga* nad dotychczas stosowanymi polega na tym, że nowa procedura ma większą moc, o czym przekonują wyniki symulacyjne z pracy Benjaminiego i Hochberga (1995). Znaczne ograniczenie tej procedury stanowi założenie o niezależności statystyk testowych. Owo założenie zostało później nieco osłabione w pracy Benjaminiego i Yekutieli (2001), w której to pracy dopuszczono pewne modele zależności.

Warto dodać, że dla procedury *Benjaminiego i Hochberga* Benjamini i Yekutieli (2001) udowodnili twierdzenie o kontroli *FDR* na poziomie α , bez czynienia jakichkolwiek założeń dotyczących zależności *p*-wartości odpowiednich testów. Pokazali oni mianowicie, że procedura step-up z $\alpha_i = i\alpha / \left\{ m \sum_{j=1}^m 1/j \right\}$ kontroluje *FDR* na poziomie α . Procedura ta ma jednak znacznie niższą moc od zwykłej procedury *Benjaminiego i Hochberga* (tzn. procedury step-up z $\alpha_i = i\alpha/m$).

4.3. Porównanie kontroli FDP i FDR. Zauważmy, że skoro dla dowolnej

zmiennej losowej X o wartościach z przedziału $[0, 1]$, prawdziwe są relacje:

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(X | X \leq \gamma) \mathbb{P}(X \leq \gamma) + \mathbb{E}(X | X > \gamma) \mathbb{P}(X > \gamma) \\ &\leq \gamma \mathbb{P}(X \leq \gamma) + \mathbb{P}(X > \gamma),\end{aligned}$$

zatem

$$\frac{\mathbb{E}(X) - \gamma}{1 - \gamma} \leq \mathbb{P}(X > \gamma) \leq \frac{\mathbb{E}(X)}{\gamma}.$$

Stąd, gdy procedura kontroluje FDR na poziomie α , to kontroluje ona FDP na poziomie α/γ , natomiast, jeśli procedura kontroluje FDP na poziomie α , to kontroluje ona FDR na poziomie $(1 - \gamma)\alpha + \gamma$.

5. Kontrola i estymacja $pFDR$ oraz pojęcie q -wartości. W niniejszym rozdziale omówimy zagadnienia dotyczące kontroli i estymacji miary $pFDR$ oraz przybliżymy pojęcie q -wartości.

5.1. Definicja i własności miary $pFDR$. Poniżej przedstawimy zaproponowaną w pracach Storey'a (2002, 2003) miarę o nazwie $pFDR$. W szczególności, podamy definicję tej miary oraz jej charakterystykę w terminach statystyki bayesowskiej. Miarę $pFDR$ definiujemy następująco

$$pFDR := \mathbb{E}\left(\frac{V}{R} \mid R > 0\right).$$

Jest to zatem wartość oczekiwana zmiennej losowej V/R pod warunkiem, że $R > 0$, czyli pod warunkiem, że wystąpiło co najmniej jedno odrzucenie.

Załóżmy, że testujemy jednocześnie m hipotez zerowych H_i , wobec hipotez alternatywnych H_i , $i = 1, \dots, m$. Przyjmijmy, że weryfikacja ta odbywa się na podstawie statystyk testowych T_1, T_2, \dots, T_m , przy wykorzystaniu wspólnego obszaru odrzuceń (zbioru krytycznego), który oznaczmy tu przez Γ .

Każdą hipotezę zerową H możemy utożsamić ze zmienną zero-jedynkową, przypisując jej wartość 0, gdy weryfikowana hipoteza jest prawdziwa, oraz wartość 1, gdy hipoteza ta jest fałszywa.

Niech π_0 oznacza prawdopodobieństwo *a priori* tego, że hipoteza H jest prawdziwa, tzn. $\pi_0 := \mathbb{P}(H = 0)$. Z kolei, oznaczmy przez π_1 prawdopodobieństwo *a priori* tego, że H jest fałszywa, tzn. $\pi_1 := \mathbb{P}(H = 1) = 1 - \pi_0$. Poniżej przytaczamy wynik z pracy Storey'a (2003), przedstawiający $pFDR$ jako odpowiednie prawdopodobieństwo *a posteriori* (zob. Storey (2003), Twierdzenie 1).

TWIERDZENIE E. *Załóżmy, że ciąg (T_i, H_i) , $i = 1, \dots, m$, jest ciągiem niezależnych, jednakowo rozłożonych wektorów losowych oraz, że rozkład statystyki testowej T_i pod warunkiem prawdziwości hipotezy zerowej H_i ma postać*

$$T_i \mid H_i = (1 - H_i)F_0 + H_iF_1,$$

gdzie: F_0 – rozkład statystyki testowej pod warunkiem prawdziwości hipotezy zerowej, natomiast F_1 – rozkład statystyki testowej pod warunkiem prawdziwości hipotezy alternatywnej.

Wtedy, co oczywiste, $\mathbb{P}(H_i = 0 \mid T_i \in \Gamma) = \mathbb{P}(H = 0 \mid T \in \Gamma)$ dla każdego $i = 1, \dots, m$ (tzn. wszystkie prawdopodobieństwa $\mathbb{P}(H_i = 0 \mid T_i \in \Gamma)$, $i = 1, \dots, m$, są identyczne, a stąd możemy opuścić indeks i), oraz

$$(11) \quad \begin{aligned} pFDR(\Gamma) &= \mathbb{P}(H = 0 \mid T \in \Gamma) \\ &= \frac{\pi_0 \mathbb{P}(T \in \Gamma \mid H = 0)}{\pi_0 \mathbb{P}(T \in \Gamma \mid H = 0) + \pi_1 \mathbb{P}(T \in \Gamma \mid H = 1)} \\ &= \frac{\pi_0 \mathbb{P}(T \in \Gamma \mid H = 0)}{\mathbb{P}(T \in \Gamma)}. \end{aligned}$$

Dowód Twierdzenia E pomijamy.

Teza powyższego twierdzenia jest o tyle zaskakująca, że wielkość $pFDR(\Gamma)$ nie zależy od m – liczby weryfikowanych hipotez.

5.2. Pojęcie q -wartości. Podamy teraz definicję wprowadzonego przez Storey'a pojęcia q -wartości. Stanowi ono bayesowski odpowiednik przedstawionego wcześniej terminu p -wartości i stąd też bywa często określane mianem p -wartości *a posteriori*.

Dla ustalonej wartości t , statystyki testowej T , q -wartość jest wielkością zdefiniowaną następująco:

$$q\text{-value}(t) = \inf_{\{\Gamma_\alpha: t \in \Gamma_\alpha\}} pFDR(\Gamma_\alpha),$$

gdzie $\{\Gamma_\alpha\}$ jest rodziną zbiorów odrzuceń (obszarów krytycznych) hipotezy zerowej, spełniających relację $\alpha < \beta \implies \Gamma_\alpha \subset \Gamma_\beta$.

Z przytoczonej definicji wynika, że q -wartość to najmniejsza wartość miary $pFDR$, przy której zaobserwowana wartość statystyki testowej prowadzi do odrzucenia hipotezy zerowej (dla porównania – zwykła p -wartość oznacza najmniejszy poziom istotności, przy którym zaobserwowana wartość statystyki testowej prowadzi do odrzucenia hipotezy zerowej).

Przy założeniach twierdzenia E mamy

$$q\text{-value}(t) = \inf_{\{\Gamma_\alpha: t \in \Gamma_\alpha\}} \mathbb{P}(H = 0 \mid T \in \Gamma_\alpha).$$

Storey zaproponował regułę testowania, polegającą na odrzuceniu hipotezy zerowej H , gdy

$$\widehat{\mathbb{P}}(H = 0 \mid T = t) \leq \alpha,$$

gdzie $\widehat{\mathbb{P}}(H = 0 \mid T = t)$ jest pewnym nieparametrycznym estymatorem wielkości $\mathbb{P}(H = 0 \mid T = t)$ (zob. Efron, Tibshirani, Storey i Tusher (2001)).

Przytoczona metoda testowania znalazła zastosowanie w analizie mikro-macierzy DNA, samego zaś pojęcia q -wartości użyto do celów oceny istotności poziomu ekspresji każdego z badanych genów (zob. Storey (2003), s. 7 oraz uwagi w rozdziale 5).

Zauważmy, że na mocy twierdzenia E (zob. relacja (11)), miarę $pFDR$ można, w terminie p -wartości p , przedstawić następująco

$$\begin{aligned} pFDR([0, \gamma]) &= \mathbb{P}(H = 0 \mid p \leq \gamma) \\ &= \frac{\pi_0 \mathbb{P}(p \leq \gamma \mid H = 0)}{\pi_0 \mathbb{P}(p \leq \gamma \mid H = 0) + \pi_1 \mathbb{P}(p \leq \gamma \mid H = 1)} \\ &= \frac{\pi_0 \mathbb{P}(p \leq \gamma \mid H = 0)}{\mathbb{P}(p \leq \gamma)}, \end{aligned}$$

gdzie γ oznacza wartość progową, taką, że przedział $[0, \gamma]$ jest wspólnym zbiorem odrzuceń weryfikowanych hipotez zerowych.

5.3. Algorytm estymacji miary $pFDR$. Przedstawimy teraz sposób estymacji miary $pFDR$, zaproponowany przez Storey'a (2003). Załóżmy, że decyzje o odrzuceniu bądź przyjęciu hipotezy zerowej H podejmujemy na podstawie odpowiedniej p -wartości p .

Niech, dla pewnego $\gamma \geq 0$, $[0, \gamma]$ oznacza wspólny zbiór odrzuceń hipotez zerowych. Z faktu, że przy założeniu prawdziwości hipotezy zerowej p -wartości mają rozkład jednostajny wynika, iż

$$(12) \quad pFDR([0, \gamma]) = \frac{\pi_0 \mathbb{P}(p \leq \gamma \mid H = 0)}{\mathbb{P}(p \leq \gamma)} = \frac{\pi_0 \gamma}{\mathbb{P}(p \leq \gamma)}.$$

Można zaproponować zatem następujący algorytm estymacji $pFDR$ (zob. także Storey (2002), wzór (8)).

Algorytm

- dla każdej z m hipotez zerowych H_1, \dots, H_m wyznaczamy odpowiadające im p -wartości p_1, \dots, p_m ;
- estymujemy $\pi_0 = m_0/m$ – proporcję hipotez prawdziwych – na mocy wzoru

$$\hat{\pi}_0(\lambda) = \frac{\#\{i : p_i > \lambda\}}{(1 - \lambda)m},$$

gdzie wartość parametru λ jest wybierana tak, aby zrównoważyć wpływ obciążenia i wariancji estymatora $\hat{\pi}_0$ (zob. Storey (2002), sekcja 9);

- estymujemy prawdopodobieństwo $\mathbb{P}(p \leq \gamma)$, według wzoru

$$\hat{\mathbb{P}}(p \leq \gamma) = \frac{\max\{1, \#\{i : p_i \leq \gamma\}\}}{m};$$

d) na mocy (12) oraz punktu b) wyznaczamy szukany estymator, ze wzoru

$$pFD\widehat{R}_\lambda([0, \gamma]) = \frac{\widehat{\pi}_0(\lambda)\gamma}{\widehat{\mathbb{P}}(p \leq \gamma)}.$$

Powyższy estymator wykazuje dobre własności asymptotyczne. Jest on także estymatorem największej wiarygodności (zob. Storey (2002), rozdział 6).

6. Zastosowania w analizie mikromacierzy DNA. W tym rozdziale przedstawimy koncepcję mikromacierzy DNA oraz wykorzystanie omówionych wcześniej procedur w analizie danych mikromacierzowych, która to analiza znajduje coraz powszechniejsze zastosowanie w genetyce i diagnostyce medycznej.

6.1. Mikromacierze DNA – opis zagadnienia. Funkcjonowanie każdego z genów można badać za pomocą pomiaru ilości odpowiadającego mu *mRNA*, będącego śladem każdego aktywnego genu, który uległ procesowi transkrypcji (czyli zachodzącemu w jądrze komórkowym procesowi przepisanania informacji genetycznej z DNA na *mRNA*). Ilość cząstek *mRNA* odpowiadających danemu genowi stanowi miarę aktywności tegoż genu. W celu badania roli i znaczenia genów opracowano koncepcję tzw. mikromacierzy DNA (*DNA microarray*). Mikromacierz DNA to płytka z regularnie naniesionymi tysiącami pól mikroskopijnej wielkości, na których można umieścić fragmenty różnych sekwencji DNA (w formie pojedynczych nici).

Standardowe techniki biochemiczne umożliwiają znakowanie *mRNA*, pochodzących z badanych tkanek, specjalnymi markerami fluorescencyjnymi w promieniach UV. Na mikromacierz DNA nakładamy próbkę cząstek *mRNA*. W wyniku tego, poszczególne *mRNA* wiążą się z polem, gdzie znajduje się DNA, z którego owe *mRNA* powstały w procesie transkrypcji. Innymi słowy, cząstki *mRNA* pochodzące z ustalonego genu wiążą się tylko w tym polu, gdzie występuje ten ustalony gen. Aktywność każdego z genów oceniamy po stopniu natężenia fluorescencji pola, w którym gen ten jest obecny. Bardzo intensywne świecenie pola wskazuje, że przyłączyło się do niego bardzo dużo cząstek *mRNA*. Oznacza to tym samym, że występujący w owym polu gen jest bardzo aktywny. Przy użyciu opisanej techniki można zidentyfikować geny aktywne w badanej tkance.

Przedstawiona koncepcja znajduje coraz liczniejsze zastosowania w diagnozowaniu i leczeniu wielu schorzeń. Umożliwia ona bowiem wykrywanie różnic między zestawem genów aktywnych w tkankach zdrowych i tkankach chorych. W szczególności, geny z komórek różnego typu nowotworów wykazują istotne różnice w poziomach ekspresji. Fakt ten odgrywa zasadniczą rolę przy wyborze metody leczenia nowotworów i jednocześnie pozwala na poszerzenie listy genów, odpowiedzialnych za powstawanie oraz rozwój tych chorób.

Dzięki znajomości różnic między zestawem genów aktywnych w zdrowej i chorej tkance można dociekać przyczyn konkretnych nowotworów. Choroba w takim ujęciu traktowana jest jako funkcja aktywności lub braku aktywności konkretnych genów. Gdy da się już ustalić, które geny funkcjonują w sposób niewłaściwy, można wówczas „zaatakować” wyłącznie te wybrane geny. Taka forma terapii molekularnej działa odmiennie od dotychczas stosowanych radio- i chemioterapii, które uszkadzają zarówno chore, jak i zdrowe komórki. Warto w tym miejscu wspomnieć również o innym sposobie wykorzystania techniki mikromacierzy DNA. Realizacja tej techniki przebiega w sposób następujący. Na początku, DNA z tkanki nowotworowej znakowany jest czerwonym barwnikiem, a ze zdrowej zielonym. Następnie, mieszaninę obu próbek DNA nakłada się na mikromacierz DNA. Podobnie jak *mRNA* we wcześniejszym opisie – oznakowane dwoma różnymi kolorami DNA ulegają przyłączeniu do odpowiednich pól mikromacierzy. W wyniku tych przyłączeń, pola zawierające geny, które uległy aktywacji w komórkach rakowych wybarwiają się na czerwono (ponieważ przyłącza się do nich więcej DNA oznakowanych czerwonym barwnikiem), natomiast pola z genami, które w komórkach rakowych są nieobecne, zabarwiają się na zielono (brakuje tam bowiem cząstek DNA z czerwonym barwnikiem, a występują jedynie te DNA, które oznakowano na zielono).

Wszystkie osoby zainteresowane tematyką i historią badań nad DNA zachęcamy do lektury popularnonaukowej książki Watsona i Berry’ego (2005). Pierwszy z autorów jest współodkrywcą struktury DNA, za co został w 1962 roku uhonorowany Nagrodą Nobla z dziedziny fizjologii i medycyny.

W naszych rozważaniach dotyczących mikromacierzy DNA skoncentrujemy się na kwestiach badania znaczenia różnic w poziomach ekspresji genów:

- 1) z tkanek osób zdrowych oraz tkanek osób dotkniętych nowotworem,
- 2) z tkanek osób chorych na dwa różne typy nowotworów.

Po wstępnym opracowaniu danych mikromacierzowych następuje dalsza analiza statystyczna, której celem jest ustalenie istotności różnic poziomów ekspresji genów dla rozważanych grup. Podstawy tej analizy w jej bardziej formalnym ujęciu przedstawimy w podrozdziale 6.2.

6.2. Porównywanie poziomów ekspresji genów – ogólna koncepcja analizy statystycznej. Niech $\mathbb{X}_j^{(1)} = (X_{1j}^{(1)}, \dots, X_{mj}^{(1)})$, $1 \leq j \leq n_1$, oznaczają niezależne, jednakowo rozłożone wektory losowe obserwacji poziomów ekspresji m genów z tkanek n_1 osób należących do pierwszej z badanych grup, natomiast $\mathbb{X}_j^{(2)} = (X_{1j}^{(2)}, \dots, X_{mj}^{(2)})$, $1 \leq j \leq n_2$, oznaczają niezależne, jednakowo rozłożone wektory losowe obserwacji poziomów ekspresji m genów pochodzących z tkanek n_2 osób należących do drugiej z badanych grup. Niech ponadto,

$X_i^{(1)}$, $1 \leq i \leq m$, będzie losowym poziomem ekspresji i -tego genu w pierwszej badanej grupie, natomiast $X_i^{(2)}$, $1 \leq i \leq m$, oznacza losowy poziom ekspresji i -tego genu w drugiej z badanych grup. Przyjmujemy, że i -ty gen ma różne poziomy ekspresji w rozważanych grupach, jeśli rozkład zmiennej losowej $X_i^{(1)}$ jest różny od rozkładu zmiennej losowej $X_i^{(2)}$. Zazwyczaj m jest znacznie większe zarówno od n_1 , jak i od n_2 , co czyni dalszą analizę problemu niezwykle skomplikowaną.

Podstawowym celem statystycznej analizy rozważanego zagadnienia jest jednoczesna weryfikacja hipotez zerowych H_i : poziomy ekspresji i -tego genu nie różnią się istotnie w obu badanych grupach, wobec hipotez alternatywnych H'_i : poziomy ekspresji i -tego genu są istotnie różne w obu badanych grupach. Przy weryfikacji tej wykorzystuje się statystyki T_i , takie jak: zwykła statystyka t -Studenta, F -statystyka czy też statystyka Wilcozona.

Ze względu na to, że rozkład statystyki T_i , przy założeniu prawdziwości hipotezy H_i , jest w praktyce nieznan, musimy estymować odpowiednie p -wartości metodą permutacyjną lub metodą bootstrap. Dopasowane p -wartości (*adjusted p-values*), będące narzędziami, którymi często operujemy przy korzystaniu z procedur jednoczesnego testowania wielu hipotez, estymujemy zastępując w odpowiedniej formule zwykle p -wartości ich estymatorami.

6.3. Wykorzystanie procedur single-step kontroli FWER do analizy danych dotyczących mikroRNA. Interesujący przykład zastosowania rozważanej przez nas tematyki jednoczesnego testowania wielu hipotez przedstawiono w pracy Lu i współautorów (2005). W artykule tym wykorzystano procedurę *Bonferroniego*, kontroli miary *FWER*, do badań dotyczących klasy RNA o nazwie *mikroRNA*, w skrócie – *miRNA*. Owe *miRNA* stały się w ostatnich latach przedmiotem intensywnych studiów i analiz. Wiąże się to ze znaczącą rolą, jaką *miRNA* odgrywają w procesach translacji *mRNA*, tzn. w procesach biosyntezy białek, które to procesy zachodzą w oparciu o informacje zakodowane w *mRNA*. Jako czynniki transkrypcji białek, *miRNA* wywierają duży wpływ na obfitość występowania wielu różnych rodzajów białek w komórce. We wspomnianym artykule Lu i współpracownicy analizowali poziomy *miRNA* w komórkach pochodzących z tkanek zajętych nowotworem oraz w komórkach pochodzących z tkanek bez obecności nowotworu.

Uzyskane rezultaty były dosyć zdumiewające. Okazało się bowiem, iż analizy polegające na obserwacji wielu *miRNA* w komórkach ssaków pozwalają na lepsze rozpoznawanie, diagnozowanie i różnicowanie stanów nowotworowych oraz źródeł ich pochodzenia, niż podobne analizy oparte na obserwacji poziomów ekspresji zwykłych *mRNA*, mierzonych dla tych samych komórek. Dokładniej, Lu i współautorzy porównywali poziomy eks-

presji znanych ludzkich *miRNA* w komórkach z 46 rakowych i 140 zdrowych tkanek – łącznie rozważano zatem informacje genetyczne pochodzące ze 186 tkanek. Dane dla 186 elementów próby opisano przy użyciu zero-jedynkowych zmiennych objaśnianych Y_i (1 dla tkanek nowotworowych, 0 dla tkanek zdrowych) i $J = 155$ -wymiarowych wektorów zmiennych objaśniających $X_i = (X_i(j) : j = 1, \dots, J)$, określających miary ekspresji dla każdego z $J = 155$ *miRNA*, gdzie $i = 1, \dots, 186$ oznaczają numery poszczególnych elementów próby (w tym przypadku tkanek).

W pracy Lu i współautorów (2005) przeprowadzono analizy porównawcze dotyczące miar ekspresji *miRNA* dla tkanek rakowych i tkanek wolnych od nowotworów. Użyto do tego procedury *Bonferroniego* ze statystyką testową *t*-Studenta, stosowaną przy weryfikacji hipotez o równości średnich (średnich poziomów ekspresji) w dwóch populacjach. Przy kontroli miary *FWER* na poziomie $\alpha = 0,05$ otrzymano, że w 59% przypadków poziomy ekspresji *miRNA* były istotnie niższe w tkankach nowotworowych niż zdrowych, a jedynie w przypadku kilku tylko tkanek rakowych poziomy ekspresji *miRNA* zostały, i to bardzo nieznacznie, przekroczone, w stosunku do poziomów ekspresji *miRNA* w tkankach wolnych od nowotworu.

Alternatywnym podejściem, stosowanym do identyfikacji różnic w poziomach ekspresji *miRNA* w tkankach nowotworowych i tkankach zdrowych, jest dopasowanie, dla każdego *miRNA*, modelu regresji logistycznej, uwzględniającego rodzaj tkanki jako zmiennej objaśniającej. W szczególności, model regresji logistycznej dla ustalonego j -tego *miRNA* jest opisany relacją

$$(13) \text{logit} \{E(Y|X(j))\} \equiv \alpha(j) + \beta(j)X(j) + \gamma(j)W(j), \quad j = 1, \dots, J,$$

gdzie: $\text{logit}(z) := \log(z/(1-z))$, $W(j)$ oznacza 19-wymiarowy wektor o współrzędnych 0 lub 1, identyfikujący typ tkanki, a $\gamma(j)$ jest (również 19-wymiarowym) wektorem parametrów, odpowiadającym wektorowi $W(j)$, przy czym 19 oznacza tu liczbę różnych typów tkanek, występujących w opracowanym przez Lu i współautorów zbiorze danych (dotyczącym np. tkanek żołądka, tkanek trzustki itd., itp.).

Parametrem stanowiącym przedmiot zainteresowania jest współczynnik $\beta(j)$ w (13), związany z miarą ekspresji $X(j)$, j -tego *miRNA*, $j = 1, \dots, J$ ($J = 155$). Dla każdego *miRNA* rozważamy dwustronne testy hipotez zerowych $H_j: \beta(j) = 0$, tzn. hipotez, że nie ma związku między poziomem ekspresji $X(j)$, j -tego *miRNA*, a występowaniem nowotworów w badanych tkankach, wobec hipotez alternatywnych $H'_j: \beta(j) \neq 0$, o istnieniu takowego związku. Statystyki testowe do weryfikacji hipotez H_j mają wtedy postać

$$T_n(j) = \frac{\beta_n(j)}{\sigma_n(j)},$$

gdzie: n – liczba tkanek, które są poddawane badaniu, $\beta_n(j)$ – estymator parametru $\beta(j)$ w modelu regresji logistycznej (13), $\sigma_n(j)$ – standardowy błąd estymatora $\beta_n(j)$.

W celu dokonania jednoczesnej weryfikacji hipotez H_j Pollard i współautorzy (2005) opracowali specjalną procedurę typu single-step, kontroli $FWER$, będącą pewną modyfikacją procedury $maxT$. Jej zastosowanie przyniosło podobne rezultaty do tych, jakie uzyskał Lu wraz ze współpracownikami. Okazało się bowiem, że w przypadku 102 hipotez, stanowiących 66% spośród wszystkich 155 hipotez dotyczących poziomów ekspresji 155 przebadanych *miRNA*, odpowiednie dopasowane p -wartości dla zmodyfikowanej procedury $maxT$, oszacowane na podstawie próby bootstrapowej wielkości $B = 5000$, nie przekraczały zakładanego z góry poziomu 0,05, miary $FWER$. Wynik ten wskazywał na istotne różnice w poziomach ekspresji *miRNA* w przypadku tkanek nowotworowych i tkanek zdrowych, przy czym w przypadku tkanek rakowych – poziomy ekspresji *miRNA* były niższe od poziomów ekspresji *miRNA* w tkankach zdrowych.

Interesującym, z biologicznego punktu widzenia, zagadnieniem jest także kwestia identyfikacji tych *miRNA*, których miary ekspresji są ze sobą skorelowane. Pollard i współautorzy (2005) rozważali $m = J(J - 1)/2 = (155 \cdot 154)/2 = 11\,935$ różnych współczynników korelacji Pearsona między parami miar ekspresji *miRNA*:

$$\rho(j, k) := \text{Corr}(X(j), X(k)), \quad j = 1, \dots, J - 1, k = j + 1, \dots, J.$$

Celem ich analiz była identyfikacja wszystkich tych par *miRNA*, w przypadku których korelacja ich poziomów ekspresji jest statystycznie istotna. Dokładniej, dla każdej pary (j, k) , różnych *miRNA*, rozważano dwustronne testy hipotez zerowych $H_{j,k}: \rho(j, k) = 0$ – o braku związku między poziomami ekspresji j -tego i k -tego *miRNA* w badanej próbie tkanek. Hipotezy alternatywne stanowiły z kolei hipotezy $H'_{j,k}: \rho(j, k) \neq 0$ – o istnieniu zależności między miarami ekspresji odpowiednich *miRNA*. Hipotezy zerowe $H_{j,k}$ weryfikowano na podstawie statystyk testowych postaci

$$T_n(j, k) = \sqrt{n}\rho_n(j, k), \quad j = 1, \dots, J - 1, k = j + 1, \dots, J,$$

gdzie: n – liczba badanych tkanek, $\rho_n(j, k)$ – odpowiednio określone empiryczne współczynniki korelacji.

W celu jednoczesnej weryfikacji $m = 11\,935$ hipotez $H_{j,k}$, o braku związku między miarami ekspresji ustalonych par *miRNA* w badanych tkankach, Pollard i współautorzy (2005) zastosowali wspomnianą wcześniej modyfikację procedury $maxT$, kontroli miary $FWER$ (dla próby bootstrapowej wielkości $B = 5000$). Okazało się, że przy nominalnym poziomie $FWER \leq \alpha = 0,05$, różne od zera i statystycznie istotne wartości współczynników korelacji były zawarte między 0,26 a 0,99. Wynika stąd konklu-

zja, iż wiele spośród par *miRNA* o istotnych wartościach współczynników korelacji przyczynia się do powstawania lub rozwoju nowotworów.

6.4. Wykorzystanie algorytmów estymacji $pFDR$ w analizie danych mikromacierzowych. Interesującą koncepcję statystycznej analizy danych mikromacierzowych, polegającą na estymacji miary $pFDR$ (zob. rozdział 5), zaproponowano w artykule Storey'a (2001). Autor przywołuje w nim nieopublikowaną pracę Riegera (2001), w której dokonano analizy danych dotyczących poziomów ekspresji 3000 genów, poddanych jednoczesnemu działaniu promieniowania jonizacyjnego. Wspomniana analiza polegała na porównaniu wpływu tego promieniowania na wartości poziomów ekspresji genów w dwóch wydzielonych grupach pacjentów, a mianowicie – w grupie zwykłych pacjentów (grupa 1) oraz w grupie pacjentów wrażliwych na owo promieniowanie (grupa 2). Dane empiryczne przedstawiały się następująco: grupa 1 liczyła 15 osób, natomiast w grupie 2 znalazło się 13 osób. Dla 3000 genów zmierzono ich poziomy ekspresji w tkankach osób z wyróżnionych grup i obliczono wartości statystyk testowych według wzoru na wartość t -statystyki, stosowanego przy weryfikacji hipotezy o równości średnich w dwóch populacjach. Zdecydowano, iż odrzucane będą hipotezy zerowe dotyczące genów, w przypadku których wartości bezwzględne odpowiadających im statystyk testowych są większe od 2. Okazało się, że takich odrzuconych genów było 146. Na tej podstawie oszacowano $pFDR$ w sposób następujący. Dokonano losowej permutacji ciągu

$$(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2),$$

15 „jedynek” (grupa 1) oraz 13 „dwójek” (grupa 2) i, jak poprzednio – wyznaczono wartości t -statystyk oraz zliczono te z nich, których wartości bezwzględne przekraczały 2. Przeprowadzając 100 takich permutacji, otrzymano średnią wartość statystyk odrzuconych hipotez dla tych permutacji równą 12,3. W efekcie – oszacowana wartość $pFDR$ wyniosła $12,3/146 = 8,44\%$.

Okazuje się, że taki prosty estymator miary $pFDR$ jest dodatnio obciążony. W celu poprawy jego jakości Storey zaproponował pomnożenie wartości tego estymatora przez wartość estymatora $\pi_0 = m_0/m$, tzn. estymatora stosunku liczby hipotez prawdziwych do liczby wszystkich hipotez zerowych. Aby otrzymać wartość $\hat{\pi}_0$ – estymatora π_0 – wybierano wartości tych t -statystyk, których wartości bezwzględne były małe, a dokładniej – mniejsze od 0,15. Spośród zaobserwowanych wartości t -statystyk, 668 miało wartości bezwzględne ze wspomnianego zakresu. Z kolei, w przypadku danych otrzymanych dla poszczególnych permutacji, średnio 750 t -statystyk miało taką własność. W efekcie, wartość estymatora π_0 wyniosła $\hat{\pi}_0 = 668/750 \approx 0,890$. Mnożąc ją przez wcześniej oszacowaną wartość miary $pFDR$, uzyskano taką oto poprawioną, nową wartość estymatora $pFDR$: $0,890 \cdot 8,44\% = 7,52\%$.

Storey rozważał również szczególną sytuację, gdy weryfikowane hipotezy są od siebie niezależne, a statystyki testowe mają brzegowe rozkłady t -Studenta. Wtedy, jeśli np. T_{26} ma rozkład t -Studenta o 26 stopniach swobody, to $P(|T_{26}| > 2) = 0,056$. Stąd, ponieważ $3000 \cdot 0,056 = 168$, oszacowana wartość $pFDR$ wyniosła $168/146 > 1$. Okazało się zatem, że założenie, iż rozkłady brzegowe statystyk testowych są rozkładami t -Studenta może stanowić pewien problem. Jednakże, 1232 wśród $3000 \cdot 100$ wartości statystyk testowych dla 100 opisanych wcześniej permutacji miało wartości bezwzględne większe od 2, co stanowiło 0,41%. Tym samym, estymowana średnia wartość fałszywych odrzuceń hipotez zerowych wyniosła $0,0041 \cdot 3000 = 12,3$, co pokryło się z wartością uzyskaną wcześniej.

6.5. Zastosowanie algorytmu estymacji q -wartości do analizy danych mikromacierzowych. Skutecznym pomysłem, umożliwiającym ustalenie jak największej liczby genów o istotnie różnych poziomach ekspresji, okazało się użycie miary FDR i związanego z nią pojęcia q -wartości. Podejście takie jest bardziej elastyczne (w sensie mocy stosowanych procedur) w stosunku do metod opartych na rozważaniu kontroli miary $FWER$.

Jako przykład wykorzystania wspomnianego pomysłu przedstawimy wyniki eksperymentu, w którym badano różnice w poziomach ekspresji genów dla tkanek zaatakowanych przez dwa różne rodzaje nowotworu płuc: $BRCA1$ oraz $BRCA2$ (zob. Storey i Tibshirani (2003)). Dane po wstępnej selekcji obejmowały 3170 genów dla 7-elementowej grupy chorych na $BRCA1$ ($n_1 = 7$) oraz 8-elementowej grupy chorych na $BRCA2$ ($n_2 = 8$). Korzystając z testu t -Studenta dla rozważanych dwóch populacji o niejednorodnych wariancjach, testowano istotność różnic w poziomach ekspresji dla każdego spośród 3170 genów. Dokładniej, weryfikowano hipotezy zerowe H_i postaci $H_i: \mu_{i1} = \mu_{i2}$, wobec hipotez alternatywnych H'_i postaci $H'_i: \mu_{i1} \neq \mu_{i2}$, $i = 1, \dots, 3170$, gdzie μ_{i1} (μ_{i2}) oznaczają średnie poziomy ekspresji w pierwszej (drugiej) z badanych populacji. Po transformacji wartości tych poziomów za pomocą funkcji $\log_2(\cdot)$, zastosowano statystykę testową

$$T_i = \frac{\bar{x}_{i2} - \bar{x}_{i1}}{\sqrt{\frac{s_{i1}^2}{n_1} + \frac{s_{i2}^2}{n_2}}},$$

gdzie \bar{x}_{i1} , s_{i1}^2 (\bar{x}_{i2} , s_{i2}^2) – próbkowe: średnia i wariancja dla pierwszej (drugiej) populacji, $i = 1, \dots, 3170$.

Ze względu na to, że rozkłady statystyk T_i , przy założeniu prawdziwości hipotez zerowych H_i , są nieznanne, odpowiednie p -wartości oszacowano za pomocą metody permutacyjnej, którą można opisać w sposób następujący.

Oznaczmy przez (t_1^b, \dots, t_s^b) , $b = 1, \dots, B$, jedną z $B = 100$ permutacji ciągu (t_1, \dots, t_s) , realizacji statystyk T_i , $i = 1, \dots, s$, $s = 3170$. Wtedy odpo-

wiednia p -wartość jest dana wzorem

$$p_i = \sum_{b=1}^B \frac{\#\{j : |t_j^b| \geq |t_i|, j = 1, \dots, s\}}{sB}.$$

Do oceny istotności różnic poziomów ekspresji badanych genów zastosowano odpowiedni algorytm estymacji q -wartości (zob. Storey i Tibshirani (2003), s. 9445). W efekcie, otrzymano, że w przypadku 160 genów – odpowiednie q -wartości spełniały warunek $q \leq 0,05$. Należy zatem oczekiwać, że około 8 spośród owych 160 genów zaklasyfikowanych do klasy genów, których poziomy ekspresji różnią się istotnie w obu badanych grupach (populacjach) – zostanie błędnie do tej klasy przypisanych. Dla porównania, przy zastosowaniu algorytmu opartego na tradycyjnej procedurze, polegającej na sprawdzaniu, czy zwykle p -wartości przekraczają ustalony próg (*cut-off value*) – okazało się, że genów o istotnie różnych poziomach ekspresji w badanych grupach było 51 – gdy p -wartości nie przekraczały 0,001, oraz między 9 a 11 – gdy p -wartości nie przekraczały 0,0001.

7. Uwagi końcowe. Pierwsze badania dotyczące zagadnień jednoczesnego testowania wielu hipotez statystycznych wykorzystywały metody porównań wielokrotnych.

Niezwykle dynamiczny w ostatnich latach rozwój nauk przyrodniczych, w szczególności zaś medycyny i biologii molekularnej, wymusił konieczność opracowania nowych procedur statystycznej weryfikacji wielu hipotez. Mimo bezsprzecznych sukcesów na tym polu, nie w pełni powiodły się podejmowane dotychczas próby stworzenia spójnej teorii dotyczącej wspomnianych zagadnień, podobnej do tej, jaką opracowano w przypadku testowania pojedynczych hipotez statystycznych. Pewne, przynajmniej, że dość obiecujące, próby w tym kierunku poczyniono w pracy Lehmana, Romano i Shaffer (2005). Autorzy opisali w niej zagadnienia jednoczesnej weryfikacji wielu hipotez przy użyciu terminów z teorii decyzji statystycznych. Główny wynik wspomnianej pracy stanowi dowód tego, że wśród pewnych reguł decyzyjnych, będących procedurami step-down lub step-up kontroli *FWER*, można wskazać te, które maksymalizują prawdopodobieństwo słusznego odrzucenia hipotez fałszywych.

Literatura

- [1] D. B. Allison, G. L. Gadbury, M. Heo, J. R. Fernandez, C. K. Lee, T. A. Prolla, R. Weindruch, (2002), *A mixture approach for the analysis of microarray gene expression data*, Comput. Statist. Data Anal., 39, 1–20.
- [2] Y. Benjamini, Y. Hochberg, (1995), *Controlling the false discovery rate: a practical and powerful approach to multiple testing*, J. Roy. Statist. Soc. Ser. B 57, 289–300.

- [3] Y. Benjamini, D. Yekutieli, (2001), *The control of the false discovery rate in multiple testing under dependency*, Ann. Statist. 29, 1165–1188.
- [4] S. Dudoit, J. P. Shaffer, J. C. Boldrick, (2003), *Multiple hypothesis testing in microarray experiments*, Statistical Science, 18 (1), 71–103.
- [5] B. Efron, R. Tibshirani, J. D. Storey, V. Tusher, (2001), *Empirical Bayes analysis of a microarray experiment*, J. Amer. Statist. Assoc. 96, 1151–1160.
- [6] J. A. Ferreira, A. H. Zwinderman, (2006), *On the Benjamini-Hochberg method*, Ann. Statist. 34, 1827–1849.
- [7] A. Gelman, (2003), *A Bayesian formulation of explanatory data analysis and goodness-of-fit testing*, International Statistical Review 71, 369–382.
- [8] C. Genovese, L. Wasserman, (2002), *Operating characteristics and extensions of the false discovery rate procedure*, J. Roy. Statist. Soc. Ser. B Stat. Methodol. 64, 499–518.
- [9] C. Genovese, L. Wasserman, (2004), *A stochastic process approach to false discovery control*, Ann. Statist. 32, 1035–1061.
- [10] Z. Guan, B. Wu, H. Zhao, (2004), *Model-based to FDR estimation*, Submitted.
- [11] Y. Hochberg, (1988), *A sharper Bonferroni procedure for multiple tests of significance*, Biometrika 75, 800–802.
- [12] S. Holm, (1979), *A simple sequentially rejective multiple test procedure*, Scand. Journal Statist. 6, 65–70.
- [13] G. Hommel, (1986), *Multiple test procedures for arbitrary dependence structures*, Metrika 33, 321–336.
- [14] G. Hommel, T. Hoffman, (1988), *Controlled uncertainty*, in: *Multiple Hypotheses Testing* (P. Bauer, G. Hommel and E. Sonnemann, eds.), 154–161. Springer, Heidelberg.
- [15] R. Hubbard, M. J. Bayarri, (2003), *P-values are not error probabilities*. Available in Internet.
- [16] V. Iyer, S. Sarkar, (2007), *An adaptive single-step FDR procedure with applications to DNA microarray analysis*, Biometrical Journal 49, 127–135.
- [17] J. Koronacki, J. Mielniczuk, (2004), *Statystyka dla studentów kierunków technicznych i przyrodniczych*, WNT, Warszawa.
- [18] E. L. Lehmann, (1957), *A theory of some multiple decision problems*, I. Ann. Statist. 28, 1–25.
- [19] E. L. Lehmann, J. P. Romano, (2005), *Generalizations of the familywise error rate*, Ann. Statist. 33, 1138–1154.
- [20] E. L. Lehmann, J. P. Romano, (2005), *Testing Statistical Hypotheses*, 3rd ed., Springer, New York.
- [21] E. L. Lehmann, J. P. Romano, J. Shaffer, (2005), *On optimality of stepdown and stepup multiple test procedures*, Ann. Statist. 33, 1084–1108.
- [22] K. S. Pollard, M. J. van der Laan, M. D. Birkner, S. Dudoit, (2005), *Test statistics null distributions in multiple testing: simulation studies and applications to genomics*, U. C. Berkeley Division of Biostatistics Working Paper Series, paper 184.
- [23] J. P. Romano, M. Wolf, (2005), *Exact and approximate stepdown methods for multiple hypothesis testing*, J. Amer. Statist. Assoc. 100, 94–108.
- [24] J. P. Romano, A. M. Shaikh, (2006), *Stepup procedures for Control of Generalizations of the Familywise Error Rate*, Ann. Statist. 34, 1850–1873.
- [25] S. Sarkar, C. Chang, (1997), *The Simes method for multiple hypothesis testing with positively dependent test statistics*, J. Amer. Statist. Assoc. 92, 1601–1608.
- [26] R. Simes, (1986), *An improved Bonferroni procedure for multiple tests of significance*, Biometrika 73, 751–754.

- [27] J. Storey, (2002), *A direct approach to false discovery rates*, J. Roy. Statist. Soc. Ser. B Stat. Methodol. 64, 479–498.
- [28] J. Storey, (2003), *The positive false discovery rate: A Bayesian interpretation and the q -value*, Ann. Statist. 31, 2013–2035.
- [29] J. Storey, R. Tibshirani, (2003), *Statistical significance for genomewide studies*, PNAS, 1000, 9440–9445.
- [30] A. Victor, G. Hommel, (2007), *Combining adaptive designs with control of the false discovery rate – a generalized definition for a global p -value*, Biometrical Journal 49, 94–106.
- [31] J. D. Watson, A. Berry, (2005), *DNA. Tajemnica życia*, wyd. CiS oraz W. A. B., Warszawa.
- [32] P. H. Westfall, S. S. Young, (1993), *Resampling-based multiple testing: Examples and methods for p -value adjustment*, John Wiley & Sons.

Marcin Dudziński
Katedra Ekonometrii i Statystyki SGGW
ul. Nowoursynowska 166
02-787 Warszawa, Poland
E-mail: mdudzinski@mors.sggw.waw.pl

Konrad Furmańczyk
Katedra Zastosowań Matematyki SGGW
ul. Nowoursynowska 166
02-787 Warszawa, Poland
E-mail: konfur@wp.pl

Simultaneous testing procedures of many null hypotheses and their applications in the DNA microarrays analysis

Abstract. In our paper, we consider different approaches to the problem of simultaneous testing of many null hypotheses. In this context, we discuss the single-step, the step-down and the step-up procedures of multiple testing. In particular, we are concerned with their properties and applications in the control of the error rates, such as: FWER, k -FWER, FDP, FDR, pFDR. The mentioned procedures are intensively used in the DNA microarrays analysis, which enables the monitoring of expression levels of many genes simultaneously and is widely applied in recent medical diagnostics, treatment and research.

Key words: multiple hypothesis testing procedures, familywise error rate, false discovery proportion, false discovery rate, p -values, q -values, DNA microarrays.

(wpłynęło 10 grudnia 2006 r.)