

Uniwersytet Warszawski
Wydział Matematyki, Informatyki i Mechaniki

Marcin Kosiński

Nr albumu: 291505

**Testowanie zbioru hipotez na
przykładzie danych o leczeniu
pacjentów dotkniętych zawałem**

**Praca licencjacka
na kierunku MATEMATYKA**

Praca wykonana pod kierunkiem
dra Błażeja Miasojedowa
Instytut Matematyki Stosowanej i Mechaniki
Zakład Statystyki Matematycznej

Czerwiec 2013

Oświadczenie kierującego pracą

Potwierdzam, że niniejsza praca została przygotowana pod moim kierunkiem i kwalifikuje się do przedstawienia jej w postępowaniu o nadanie tytułu zawodowego.

Data

Podpis kierującego pracą

Oświadczenie autora (autorów) pracy

Świadom odpowiedzialności prawnej oświadczam, że niniejsza praca dyplomowa została napisana przeze mnie samodzielnie i nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami.

Oświadczam również, że przedstawiona praca nie była wcześniej przedmiotem procedur związanych z uzyskaniem tytułu zawodowego w wyższej uczelni.

Oświadczam ponadto, że niniejsza wersja pracy jest identyczna z załączoną wersją elektroniczną.

Data

Podpis autora (autorów) pracy

Streszczenie

Praca przedstawia teorię i zastosowania zagadnienia testowania zbioru hipotez. Praca została podzielona na 3 części. Pierwsza to teoretyczny opis zagadnienia i działania 3 procedur kontroli miary błędu FWER: Bonferroniego, Holma i Hochberga. Druga część pracy skupiona została na przedstawieniu funkcji pakietu statystycznego \mathcal{R} , zastosowanych podczas analizy danych medycznych w części trzeciej, w celu porównania skuteczności dwóch metod przeprowadzania zabiegu w razie wystąpienia zagrożenia życia jakim jest zawał. Do analizy danych wykorzystano nieparametryczny test Manna-Whitneya-Wilcoxona. Dane pochodzą z I Oddziału Kardiologii Inwazyjnej Warszawskiego Uniwersytetu Medycznego.

Słowa kluczowe

testowanie zbioru hipotez, współczynnik FWER, procedura Hochberga, procedura Holma, procedura Bonferroniego, pakiet \mathcal{R} , dopasowana p-wartość, test Manna-Whitneya-Wilcoxona

Dziedzina pracy (kody wg programu Socrates-Erasmus)

11.2 Statystyka

Klasyfikacja tematyczna

62F03 Hypothesis testing

Tytuł pracy w języku angielskim

Multiple hypothesis testing with an application to data on the treatment of patients suffering from myocardial infarction.

Spis treści

Wprowadzenie	9
1. Teoria	11
1.1. Testowanie hipotez statystycznych	11
1.1.1. Model statystyczny	11
1.1.2. Weryfikacja hipotez	12
1.1.3. Błędy I i II rodzaju	13
1.1.4. Statystyki testowe	14
1.1.5. Wartość krytyczna	14
1.2. Testowanie zbioru hipotez	15
1.2.1. Zagadnienie testowania zbioru hipotez	15
1.2.2. Problemy testowania zbioru hipotez	16
1.2.3. Współczynniki miary błędu	17
1.2.4. Kontrola współczynników miary błędu	17
1.2.5. Dopasowana p-wartość	18
1.2.6. Klasyczne procedury testowania zbioru hipotez	18
1.3. Test kombinatoryczny Manna-Whitneya-Wilcoxona	23
1.4. Współczynnik korelacji Pearsona	24
2. Funkcje do testowania zbioru hipotez w pakiecie \mathcal{R}	25
2.1. Pakiet <code>stats</code>	25
2.1.1. Funkcja <code>wilcox.test()</code>	25
2.1.2. Obiekt klasy "htest"	26
2.1.3. Funkcja <code>cor()</code> i funkcja <code>pairs()</code>	27
2.2. Pakiet <code>multtest</code>	28
2.2.1. Funkcja <code>mt.rawp2adjp()</code>	28
2.2.2. Funkcja <code>mt.reject()</code>	29
2.2.3. Funkcja <code>mt.plot()</code>	29
2.2.4. Porównanie procedur Hochberga i Holma przy użyciu pakietu <code>multtest</code>	30
3. Analiza danych medycznych	33
3.1. Opis danych	33
3.2. Analiza korelacji	35
3.3. Test kombinatoryczny Manna-Whitneya-Wilcoxona	36
3.4. Testowanie zbioru hipotez testem Manna-Whitneya-Wilcoxona	38
3.5. Wyniki i wnioski	40
Podsumowanie	41

A. Kody pakietu \mathcal{R} użyte w pracy	43
Bibliografia	47

Spis rysunków

1.1. Prawdopodobieństwo popełnienia co najmniej jednego błędu I rodzaju dla różnych poziomów istotności α i liczby hipotez m	16
1.2. Schemat procedury typu step-down.	22
1.3. Schemat procedury typu step-up.	22
2.1. Poziom kontroli miary FWER, a liczba odrzuconych hipotez - przykład	31
3.1. Macierz korelacji i wykresy rozrzutu analizowanych parametrów serca.	35
3.2. Wykresy pudełkowe analizowanych parametrów serca	37
3.3. Poziom kontroli miary FWER, a liczba odrzuconych hipotez - analiza	39

Spis tabel

1.1. Możliwe scenariusze w procesie testowania hipotezy zerowej.	13
2.1. Opis argumentów funkcji <code>wilcox.test</code>	26
2.2. Opis wybranych wartości obiektu klasy <code>"htest"</code>	26
2.3. Opis argumentów funkcji <code>cor</code>	27
2.4. Opis argumentów funkcji <code>pairs</code>	27
2.5. Opis argumentów funkcji <code>mt.rawp2adjp</code>	28
2.6. Opis argumentów funkcji <code>mt.reject</code>	29
2.7. Opis argumentów funkcji <code>mt.plot</code>	29
3.1. Wyniki przeprowadzonych testów Manna-Whitneya-Wilcoxona	36
3.2. Statystyki agregujące analizowanych parametrów serca.	37

Wprowadzenie

W niniejszej pracy przedstawiono zagadnienie testowania zbioru hipotez ze wskazaniem na problemy związane z kontrolą błędu I rodzaju, na które narażone są metody statystyczno-decyzyjne, w przypadku rozważania mnogiej liczby hipotez. Opisano powstałe, w związku z problemami, procedury odpowiedniego podejścia do weryfikacji postawionych hipotez. Z tym zagadnieniem można się spotkać, gdy skala zjawiska wymaga uwzględnienia dziesiątek, setek, a nawet tysięcy rozważanych hipotez, a oparta o nie jedna, generalna decyzja nie może przekraczać z góry zadanego poziomu kontroli błędu I rodzaju. W przeciągu ostatnich kilkudziesięciu lat, zagadnienie testowania zbioru hipotez znalazło coraz śmielsze zastosowania i jest wykorzystywane z coraz większą popularnością w biologii molekularnej i jej badaniach przesiewowych dotyczących ekspresji genów, czy we wstępnych wysoko-przepustowych badaniach skuteczności nowych leków [4]. Korekta decyzyjna oparta na zagadnieniu testowania zbioru hipotez jest również wykorzystywana w analizie mikromacierzy DNA, która jest ostatnio szeroko stosowana w diagnostyce, leczeniu i badaniach medycznych. Liczne aplikacje teorii testowania zbioru hipotez można również znaleźć w badaniach dotyczących biosyntezy białek [7].

W pracy użyto klasycznych procedur testowania: Bonferroniego, Holma i Hochberga oraz wykorzystano możliwości jakie daje pakiet statystyczny \mathcal{R} [16], by zbadać i porównać skuteczność dwóch metod zabiegów, stosowanych wśród grupy pacjentów dotkniętych zawałem. Dane analizowane w pracy zostały zebrane przez Oddział Kardiologii Inwazyjnej Warszawskiego Uniwersytetu Medycznego. Celem pracy jest próba stwierdzenia, który z rozważanych zabiegów jest efektywniejszy.

Praca jest podzielona na trzy części. W pierwszym rozdziale zaprezentowano podstawy teorii weryfikacji hipotez statystycznych, by następnie zaznaczyć różnice w podejściu i główne potrzeby wprowadzania zagadnienia testowania zbioru hipotez, w szczególności określenia kontroli miary błędu FWER (ang. *familywise error rate*). Oprócz przedstawienia wymienionych już klasycznych procedur testowania, w rozdziale pierwszym podano opis testu Manna-Whitneya-Wilcoxona, który został zastosowany do analizy danych w rozdziale trzecim. Rozdział drugi prezentuje wybrane funkcje pakietu statystycznego \mathcal{R} oraz dodatkowego pakietu wspomagającego `multtest` [15], które zostały użyte do przeprowadzenia analizy danych. Rozdział trzeci przedstawia wspomnianą wcześniej analizę rzeczywistych danych medycznych oraz wyniki uzyskane przy pomocy przedstawionej teorii i omówionego narzędzia statystycznego.

W pracy zostały przedstawione kody pakietu \mathcal{R} użyte do analizy, których zestawienie i podsumowanie można znaleźć w Dodatku A.

Rozdział 1

Teoria

1.1. Testowanie hipotez statystycznych

Statystyka, ze względu na swoje szerokie zastosowania, jest nieodzownym elementem każdego badania. Naukowcy podczas swoich eksperymentów i badań, aby wypromować nową teorię naukową lub obalić już istniejącą, stawiają najróżniejsze tezy. Obserwacja zjawisk potwierdzających hipotezę skłania do przypuszczeń o jej słuszności, jednak pojawienie się zjawiska określanego przez daną teorię za mało prawdopodobne, przemawia za odrzuceniem weryfikowanej hipotezy i przyjęciem odmiennej. Testy statystyczne przychodzą z pomocą jako narzędzie weryfikacyjno-decyzyjne w momencie, gdy dane zjawiska i teorię można opisać i wypowiedzieć w terminach modelu probabilistycznego.

1.1.1. Model statystyczny

Aby mówić o zagadnieniu testowania hipotez statystycznych, niezbędne jest określenie modelu statystycznego, na którym oparte będą rozważania dotyczące weryfikowanej tezy naukowej. Idąc za [14] można podać następującą definicję:

Definicja 1.1. *Model statystyczny* określamy przez podanie rodziny $\{\mathbb{P}_\theta : \theta \in \Theta\}$ rozkładów prawdopodobieństwa na przestrzeni próbkowej Ω oraz zmiennej losowej $X : \Omega \rightarrow \mathcal{X}$, którą traktujemy jako obserwację. Zbiór \mathcal{X} nazywamy przestrzenią obserwacji, zaś Θ nazywamy przestrzenią parametrów.

Symbol θ jest nieznanym parametrem opisującym rozkład badanego zjawiska. Może być jednowymiarowy lub wielowymiarowy. Determinując opis zjawiska poprzez podanie parametru θ , jednoznacznie wyznaczany jest rozkład rozważanego zjawiska spośród całej rodziny rozkładów prawdopodobieństwa $\{\mathbb{P}_\theta : \theta \in \Theta\}$, co umożliwia określenie prawdziwości tezy.

Zakłada się, że przestrzeń próbkowa Ω jest wyposażona w σ -ciało \mathcal{F} . Wtedy:

Definicja 1.2. *Przestrzenią statystyczną* nazywa się trójkę $(\mathcal{X}, \mathcal{F}, \{\mathbb{P}_\theta : \theta \in \Theta\})$.

Wprowadzenie σ -ciała \mathcal{F} sprawia, że przestrzeń statystyczna staje się przestrzenią mierną, a więc można na niej określić rodzinę miar $\{\mathbb{P}_\theta : \theta \in \Theta\}$, dzięki której da się ustalić prawdopodobieństwa zajścia wszystkich zjawisk w rozważanej teorii.

1.1.2. Weryfikacja hipotez

Definicje i oznaczenia, zawarte w tym rozdziale, w większej części są oparte na [2, 14, 18].

Poprzez *hipotezę statystyczną* rozumie się wypowiedziane zdanie na temat rozkładu prawdopodobieństwa badanej próbki. Hipotezę opisującą badaną teorię, która jest testowana nazywa się *hipotezą zerową* i oznacza przez H_0 . Zdarzenia czy zjawiska, które według rozważanej hipotezy są niemożliwe, bardzo trudno jest określić przy zdefiniowanej, jak w poprzednim podrozdziale, przestrzeni statystycznej - nie przypisuje się im prawdopodobieństwa zajścia równego 0. Dlatego zdarzenia niesprzyjające hipotezie zerowej to zdarzenia, które są "bardzo mało prawdopodobne". Zaobserwowanie, podczas eksperymentu czy doświadczenia, zajścia zdarzenia, które określono mianem "bardzo mało prawdopodobnego", świadczy o nieprawdziwości przyjętej teorii czy hipotezy zerowej i usprawiedliwia przyjęcie hipotezy sprzecznej z hipotezą zerową, zwaną *hipotezą alternatywną*, oznaczaną H_1 . Proces weryfikacji hipotezy zerowej przeciwko hipotezie alternatywnej nosi nazwę *testu statystycznego*.

Test to reguła decyzyjna, która ma postać mierzalnej funkcji próbki:

$$\delta : \mathcal{X} \rightarrow \{0, 1\}.$$

$$\delta(X) = 1 \Rightarrow \text{decyzja o odrzuceniu } H_0 \text{ i przyjęciu } H_1,$$

$$\delta(X) = 0 \Rightarrow \text{decyzja o nieodrzućeniu } H_0.$$

Podział przestrzeni statystycznej, opisującej badane zjawisko, na zdarzenia zgodne z zakładaną hipotezą zerową i zdarzenia "bardzo mało prawdopodobne" wykluczające prawdziwość hipotezy zerowej i nakazujące przyjęcie hipotezy alternatywnej, narzuca wprowadzenie:

Definicja 1.3. *Obszarem krytycznym* nazywamy zbiór obserwacji K , które świadczą przeciwko postawionej hipotezie zerowej.

Innymi słowy, obszar krytyczny $K \subset \mathcal{X}$ jest dobierany tak, by zachodziło:

$$\mathbb{P}(X \in K) \leq \alpha,$$

dla zadanego z góry α zwanego *poziomem istotności*. Przyjęcie obszaru krytycznego K oraz poziomu istotności α jest ustaleniem kolejnej reguły decyzyjnej, jednoznacznie określającej czy rozważaną hipotezę zerową należy odrzucić, czy nie ma ku temu podstaw. W terminach funkcji δ obszar krytyczny jest równy $K = \{x \in \mathcal{X} : \delta(x) = 1\}$.

Poziom istotności oznacza na jaki błąd badania się zgodzono. W praktyce przyjmuje się $\alpha = 0.05$ lub $\alpha = 0.01$. Przykładowo, dla przyjętego $\alpha = 0.01$ może się okazać, że błędnie odrzucono hipotezę zerową na korzyść hipotezy alternatywnej. Takie zdarzenie może zajść z prawdopodobieństwem równym 0.01.

Podział przestrzeni statystycznej, związany z obszarem krytycznym K , prowadzi do wprowadzenia podziału przestrzeni parametrów Θ na przestrzeń Θ_0 zdarzeń sprzyjających H_0 i przestrzeń Θ_1 zdarzeń uprawniających do przyjęcia H_1 . Wzajemne wykluczanie się hipotez H_0 i H_1 oznacza, że $\Theta_0 \cap \Theta_1 = \emptyset$.

Uwaga. Należy mieć świadomość, że nieodrzućenie hipotezy zerowej na rzecz hipotezy alternatywnej wcale nie oznacza jej prawdziwości. Świadczy to jedynie o tym, że nie ma wystarczających podstaw, przy danym eksperymencie, do odrzucenia hipotezy zerowej.

1.1.3. Błędy I i II rodzaju

Ponieważ rozpatrywane obserwacje są losowe, to przy tak opisanej przestrzeni statystycznej, należy liczyć się z ewentualnymi błędami weryfikacyjnymi związanymi z zagadnieniem testowania hipotez statystycznych. Błędy są rozróżnione na dwa typy wymienione poniżej.

Definicja 1.4. *Błąd I rodzaju polega na odrzućeniu hipotezy zerowej, gdy jest prawdziwa.*

Definicja 1.5. *Błąd II rodzaju polega na nieodrzućeniu hipotezy zerowej, gdy jest fałszywa.*

Testując hipotezę zerową nie da się w tym samym czasie niwelować obu błędów z racji na ich wzajemne wykluczanie się, dlatego przyjęto, że błąd I rodzaju jest ważniejszy i przy doborze testów dba się właśnie o jego minimalizację. Są ku temu ważne powody. Pochopne odrzucenie H_0 na rzecz H_1 może okazać się fatalne w skutkach, dlatego w badaniach za hipotezę zerową przyjmuje się hipotezę, której błędne odrzucenie może mieć gorsze konsekwencje niż błędne jej przyjęcie. Chociażby przy badaniu nowych substancji wspomagających rozwój roślin, których działanie wciąż wiąże się z pewnym ryzykiem. Takie nowe preparaty zazwyczaj porównuje się z już stosowanymi innymi metodami zwiększania upraw rolnych dostępnymi w przemyśle, których działanie i efekty uboczne są znane. Na przykład, nie znane były efekty upraw roślin genetycznie modyfikowanych na zdrowie człowieka, spożywającego powstałe z nich produkty przez dłużej niż 3 miesiące. W takiej sytuacji najlepiej postawić bezpieczną hipotezę zerową H_0 : "znane w przemyśle substancje wspomagające rozwój uprawianych roślin są nie mniej skuteczne niż nowe". Błąd I rodzaju, polegający na pochopnym odrzuceniu H_0 , może narazić życie ludzi spożywających plony roślin genetycznie modyfikowanych na niebezpieczeństwo.

Poprawne decyzje statystyczne i błędy ilustruje Tabela 1.1:

Stan faktyczny	Przyjęcie H_0	Odrzucenie H_0
$H = 0$	decyzja poprawna	błąd I rodzaju
$H = 1$	błąd II rodzaju	decyzja poprawna

Tabela 1.1: Możliwe scenariusze w procesie testowania hipotezy zerowej.

Przy kontroli błędu I rodzaju wygodnie jest zdefiniować tzw. *funkcję mocy* testu.

Definicja 1.6. *Funkcją mocy testu δ , dla prawdopodobieństwa odrzucenia hipotezy zerowej $\mathbb{P}_\theta(\delta(X) = 1)$, nazywa się funkcję $1 - \beta(\theta)$, dla której zachodzi:*

$$\mathbb{P}_\theta(\delta(X) = 1) = 1 - \beta(\theta).$$

Funkcja mocy testu ma dwie interpretacje, w zależności od prawdziwości hipotez.

$\theta \in \Theta_0 \Rightarrow H_0$ jest prawdziwa, wtedy $1 - \beta(\theta)$ jest prawdopodobieństwem błędu I rodzaju.

$$1 - \beta(\theta) = \mathbb{P}_\theta(\delta(X) = 1)$$

$\theta \in \Theta_1 \Rightarrow H_1$ jest prawdziwa, wtedy $\beta(\theta)$ jest prawdopodobieństwem błędu II rodzaju.

$$\beta(\theta) = \mathbb{P}_\theta(\delta(X) = 0)$$

1.1.4. Statystyki testowe

Najczęściej spotykane w praktyce postacie testu wyglądają tak:

$$\delta(X) = \begin{cases} 1 & T(X) > c \\ 0 & T(X) \leq c \end{cases}, \quad (1.1)$$

gdzie $T(X)$ jest pewną szczególną statystyką określaną mianem *statystyki testowej*, a c liczbą nazywaną *wartością krytyczną*. Formalna definicja statystyki przedstawiona jest poniżej:

Definicja 1.7. *Statystyką, dla danej próbki $X = (X_1, X_2, \dots, X_n)$, nazywamy odwzorowanie mierzalne:*

$$T : \mathcal{X} \rightarrow \mathcal{R}.$$

Ustalenie poziomu krytycznego c oraz statystyki testowej $T(X)$ jest trzecią regułą podejmowania decyzji w procesie weryfikacji hipotez. Można ją połączyć z wcześniej wprowadzonymi regułami tak jak poniżej:

$$\begin{aligned} T(X) > c &\Rightarrow \delta(X) = 1 \Rightarrow X \in K \Rightarrow \text{odrzucaamy } H_0, \\ T(X) \leq c &\Rightarrow \delta(X) = 0 \Rightarrow X \notin K \Rightarrow \text{nie odrzucaamy } H_0. \end{aligned}$$

Posługując się funkcją testu jak w (1.1), hipotezę zerową odrzucaamy, gdy wartość naszej próbki, przy statystyce testowej $T(X)$, przekroczy pewną wartość krytyczną c .

Wszystkie 3 wymienione reguły postępowania są równoważne. Wybranie jednej z nich jest kwestią wygody. Istota testu pozostaje ta sama. Zarówno zaznaczenie zbioru krytycznego K , określenie funkcji testu δ czy ustalenie poziomu krytycznego c i statystyki testowej $T(X)$ są regułami podejmowania decyzji w zależności od wyniku obserwacji.

1.1.5. Wartość krytyczna

Ponieważ powstaje pytanie, jak dobierać statystyki testowe i jak ustalać poziom krytyczny testu, na szczególną uwagę w procesie testowania hipotez zasługuje wartość krytyczna, potocznie określana także mianem *p-wartości* (ang. *p-value*).

Definicja 1.8. *Wartość krytyczna (p-wartość) jest równa najmniejszemu poziomowi istotności, na którym dla obserwacji $x \in \mathcal{X}$ przyjmuje się hipotezę H_0 . Wartość krytyczna wyraża się wzorem:*

$$p(x) = \inf_{\theta \in \Theta_0} (1 - F_\theta(T(x))), \quad (1.2)$$

gdzie $F_\theta(T(x))$ jest dystrybuantą rozkładu statystyki testowej T związaną z rozkładem P_θ , przy prawdziwej hipotezie zerowej.

Wartość krytyczna jest zmienną losową określoną na odcinku $[0, 1]$, a ponadto zachodzi:

Stwierdzenie 1.1. *Jeśli $H_0 = \theta_0$ jest prawdziwa oraz dystrybuanta $F_{\theta_0}(T(X))$ jest ciągła, to*

$$p(X) \sim \mathcal{U}[0, 1]. \quad (1.3)$$

Dowód. [14] Zakładając, że $F_\theta^{-1}(x)$ istnieje, dla $0 < u < 1$, dystrybuenta zmiennej losowej danej wzorem (1.2) wynosi:

$$\begin{aligned}\mathbb{P}(p(X) \leq u) &= \mathbb{P}(1 - F_\theta(T(X)) \leq u) = \mathbb{P}(F_\theta(T(X)) > 1 - u) \\ &= \mathbb{P}(T(X) > F_\theta^{-1}(1 - u)) = 1 - \mathbb{P}(T(X) \leq F_\theta^{-1}(1 - u)) \\ &= 1 - F_\theta(F_\theta^{-1}(1 - u)) = u.\end{aligned}$$

□

W pracy termin wartość krytyczna będzie używany wymiennie z terminem p-wartość.

Dzięki wprowadzeniu p-wartości można się uniezależnić od wyboru przestrzeni obserwacji \mathcal{X} , rodziny rozkładów $\{\mathbb{P}_\theta : \theta \in \Theta\}$ i statystyki testowej $T(X)$. Używanie p-wartości jest popularnym i wygodnym sposobem określania obszarów przyjęcia i odrzucenia. Wyznaczenie p-wartości testu jest korzystne, gdyż zwalnia z konieczności odgórnego doboru poziomu istotności przed przeprowadzeniem testu. Badacz powinien interpretować p-wartość w ten sposób: im mniejsza uzyskana p-wartość, tym bardziej powinien być skłonny do odrzucenia H_0 na korzyść H_1 , ponieważ wtedy silniejsze powinno się stać jego przeświadczenie o fałszywości hipotezy zerowej i prawdziwości alternatywnej.

W przypadku, gdy dystrybuenta jest nieciągła, p-wartość zastępuje się *zrandomizowaną p-wartością* zachowującą własność podaną w Stwierdzeniu 1.1.

1.2. Testowanie zbioru hipotez

1.2.1. Zagadnienie testowania zbioru hipotez

Większość informacji składających się na obraz rozdziału 1.2 pochodzi z [2, 4, 5, 7, 12].

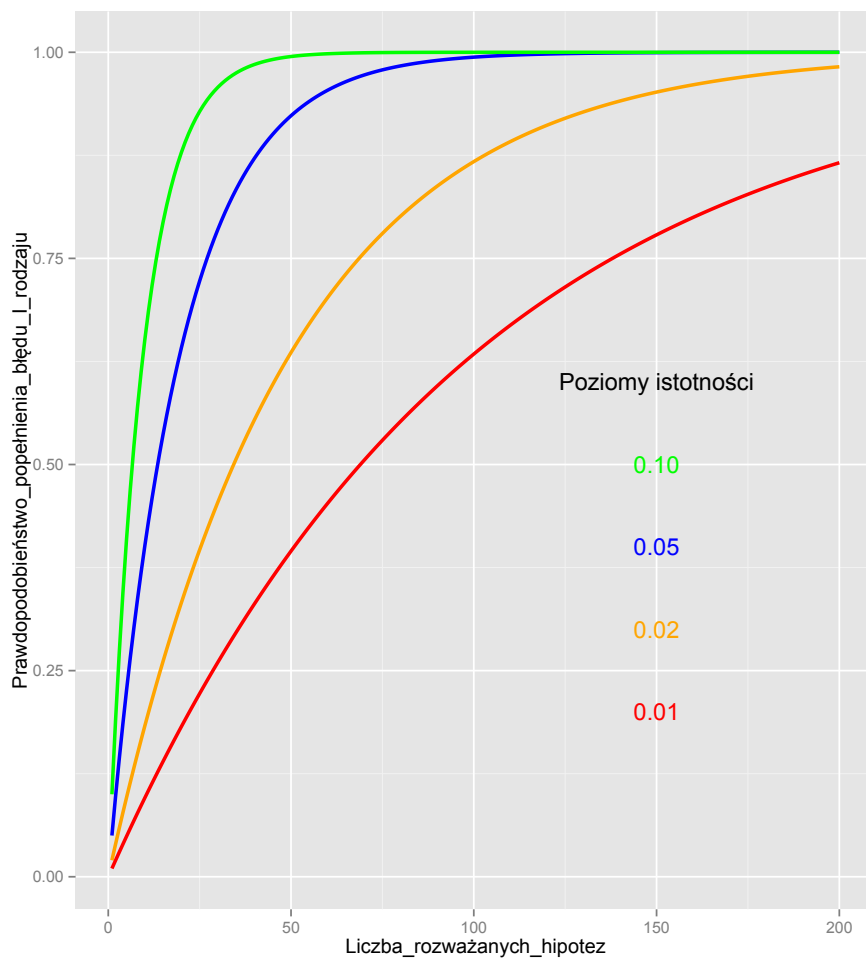
W procesie testowania teorii uwzględniającej jedynie jedną hipotezę zerową, statystyki testowe są często dobierane w ten sposób, by kontrolować błąd I rodzaju polegający na niewłaściwym odrzuceniu prawdziwej hipotezy zerowej dla zadanego z góry poziomu istotności α . Zagadnienie testowania zbioru hipotez rozważa pewną teorię, dla której weryfikacja wymaga zbadania więcej niż jednej hipotezy zerowej. Gdy testowana jest mnoga liczba hipotez m i ostateczne wnioski powinny być słuszne dla wszystkich pytań stawianych w rozważanym eksperymencie, prawdopodobieństwo błędnego odrzucenia co najmniej jednej hipotezy zerowej znacząco wzrasta wraz z liczbą stawianych hipotez m .

Jest to widoczne już w przypadku, gdy rozważany zbiór składa się z dwóch hipotez. Oznaczając te hipotezy poprzez H_0^1, H_0^2 i badając obie na poziomie istotności $\alpha = 0.05$, pod warunkiem niezależności statystyk testowych, można otrzymać, że prawdopodobieństwo iż obie hipotezy są prawdziwe wynosi $(1 - \alpha)^2 = 0.9025$. dopełnieniem tego prawdopodobieństwa, jest prawdopodobieństwo błędnego odrzucenia co najmniej jednej spośród hipotez, które wynosi $1 - (1 - \alpha)^2 = 0.0975$. Widać, że obliczone prawdopodobieństwo jest znacząco większe od zakładanego początkowo poziomu istotności $\alpha = 0.05$. W ogólności, pod warunkiem niezależności statystyk testowych, gdy testowanych jest m hipotez zerowych, prawdopodobieństwo popełnienia co najmniej jednego błędu I rodzaju wyraża się wzorem:

$$1 - (1 - \alpha)^m, \tag{1.4}$$

gdzie α to wspomniany już poziom istotności, a m to liczba rozważanych hipotez zerowych.

Na Rysunku 1.1 pokazano prawdopodobieństwo popełnienia błędu I rodzaju, dla czterech poziomów istotności $\alpha = 0.01, 0.02, 0.05$ i 0.10 i liczby hipotez $m = 1, \dots, 200$.



Rysunek 1.1: Prawdopodobieństwo popełnienia co najmniej jednego błędu I rodzaju dla różnych poziomów istotności α i liczby hipotez m . Wzrost liczby rozważanych hipotez, na ustalonym poziomie istotności α , zwiększa prawdopodobieństwo popełnienia przynajmniej jednego błędu I rodzaju.

1.2.2. Problemy testowania zbioru hipotez

Matematyczny model testowania zbioru hipotez zakłada, że weryfikując hipotezy zerowe, oznaczane H_0^i dla $i = 1, \dots, m$, dla każdej z nich dysponuje się odpowiednim testem weryfikacyjnym T_i . Aby wiedzieć, które hipotezy spośród H_0^i należy odrzucić, konstruuje się odpowiednie *procedury testowania*. Procedura testowania zbioru hipotez to reguła określająca na jakim poziomie istotności i w jakiej kolejności należy badać hipotezy zerowe. Każda procedura zakłada, że rozważane hipotezy są badane na odpowiednio dobranym, niższym poziomie istotności niż w pojedynczym przypadku. W tym celu wprowadza się pojęcie współczynników miary błędów, których używa się w zależności od kontrolowanej liczby błędów I rodzaju. Opisy niektórych współczynników i procedur znajdują się w dalszych rozdziałach tej pracy.

1.2.3. Współczynniki miary błędu

Klasyczne podejście do zagadnień testowania zbioru hipotez wyraża się poprzez konstrukcję procedury, która zapewnia kontrolę miary błędu oznaczającą prawdopodobieństwo popełnienia co najmniej jednego błędu I rodzaju. Przy V oznaczającym liczbę fałszywie odrzuconych hipotez zerowych H_0^i , miarę tę nazwano z ang. *familywise error rate* – FWER. Miara ta spełnia:

$$\text{FWER} := \mathbb{P}(V \geq 1).$$

Rozważanie podejścia kontrolującego FWER jest obarczone pewnymi wadami z zasadniczych przyczyn ujętych na Rysunku 1.1. Zdolność procedur FWER do wykrywania fałszywych hipotez zerowych maleje wraz ze wzrostem liczby stawianych hipotez zerowych. Z tego powodu postanowiono rozważać procedury kontroli miary błędu o nazwie k-FWER, które kontrolowałyby prawdopodobieństwo popełnienia co najmniej k błędów I rodzaju ($k \geq 1$). Stąd:

$$\text{k-FWER} := \mathbb{P}(V \geq k).$$

Liczba fałszywie odrzucanych hipotez nie maleje wraz ze wzrostem wszystkich odrzucanych hipotez. To stwierdzenie motywuje wprowadzenie kolejnej miary kontroli błędu zwanej z ang. *false discovery proportion* – FDP, która, dla R oznaczającego liczbę wszystkich odrzuceń, wyraża się wzorem:

$$\text{FDP} := \begin{cases} \frac{V}{R} & R > 0 \\ 0 & R = 0 \end{cases}.$$

W 1995 r. Benjamini i Hochberg [7] zaproponowali poprawkę FDP poprzez zastąpienie jej procedurą kontroli miary błędu zwaną z ang. *false discovery rate* – FDR, określanej jako wartość oczekiwana FDP. Zatem:

$$\text{FDR} := \mathbb{E}(\text{FDP}).$$

1.2.4. Kontrola współczynników miary błędu

Mówiąc o kontroli współczynników miary błędu na poziomie α , w praktyce dąży się do tego, aby zachowane były (dla pewnych $\gamma, \alpha \in (0, 1)$) nierówności:

dla FWER	dla k-FWER	dla FDP	dla FDR
$\mathbb{P}(V \geq 1) \leq \alpha$	$\mathbb{P}(V \geq k) \leq \alpha$	$\mathbb{P}(\text{FDP} > \gamma) < \alpha$	$\mathbb{E}(\text{FDP}) \leq \alpha$

Statystycy rozpatrują dwa typy kontroli miary błędu. Dla danego zagadnienia testowania zbioru hipotez H_0^i , $i = 1 \dots m$, każdą z hipotez zerowych H_0^i można utożsamić z pewnym podzbiorem przestrzeni parametrów $\Theta_i \subset \Theta$ i przedstawić w formie $H_0^i : \theta \in \Theta_i$. Niech $M_0 = \{i : H_0^i \text{ jest prawdziwa}\}$, wtedy wyróżniamy:

Definicja 1.9. Mocna kontrola miary błędu FWER na poziomie α polega na zapewnieniu:

$$\mathbb{P}_\theta(V \geq 1) \leq \alpha \text{ dla każdego } \theta \in \cap_{i \in M_0} \Theta_i.$$

Definicja 1.10. Słaba kontrola miary błędu FWER na poziomie α polega na zapewnieniu:

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

Procedura testowa zapewniająca kontrolę FWER tylko w przypadku, gdy wszystkie hipotezy zerowe są prawdziwe, zapewnia słabą kontrolę miary, zaś mocna kontrola FWER zachowywana jest wtedy, gdy jest się w stanie podać ograniczenie na FWER dla każdej możliwej konfiguracji prawdziwych i fałszywych hipotez zerowych.

1.2.5. Dopasowana p-wartość

Główną różnicą między testowaniem zbioru hipotez, a pojedynczą weryfikacją jednej hipotezy jest stosowanie dopasowanych p-wartości zamiast zwykłych p-wartości. Przy ustalonej dowolnej mierze błędu procedury testowania wielu hipotez oznaczonej przez μ , można (za [7]) podać definicję *dopasowanej p-wartości*.

Definicja 1.11. *Dopasowana p-wartość (ang. adjusted p-value) jest najmniejszą wartością rozważanej miary błędu testowania wielu hipotez, przy której hipotezę H_0^i należy odrzucić.*

Dla porównania, dopasowana p-wartość \tilde{p}_i oraz zwykła p-wartość p_i spełniają równości:

$$\tilde{p}_i = \inf\{\alpha : \text{hipoteza } H_0^i \text{ jest odrzucana na poziomie błędu } \mu = \alpha\},$$

$$p_i = \inf\{\alpha : \text{hipoteza } H_0 \text{ jest odrzucana na poziomie istotności } \alpha\}.$$

1.2.6. Klasyczne procedury testowania zbioru hipotez

W dalszej części pracy uwaga została skupiona na współczynniku kontroli miary błędu FWER. W tym podrozdziale zostaną przedstawione klasyczne procedury kontroli tego współczynnika, gdyż przeprowadzając analizę danych, postawiono za cel niedopuszczenie do popełnienia ani jednego błędu I rodzaju. Wybór FWER, zamiast wspomnianych innych współczynników, jest podyktowany tym, iż wynik badania może mieć wpływ na zdrowie wspomnianych we wprowadzeniu pacjentów dotkniętych zawałem, zatem niewskazane jest odrzucenie choć jednej prawdziwej hipotezy zerowej. Taka praktyka mogłaby być stosowana, gdy rozważanych jest wiele składników składających się na poprawność i niezawodność działania takich rzeczy jak: poduszka powietrzna, spadochron czy łódź podwodna. Produkcja tych urządzeń nie dopuszcza wyprodukowania wadliwych produktów, gdyż od ich skuteczności czy wytrzymałości zależy ludzkie życie. W przypadku, gdy sprawa ma mniejszą wagę, jak choćby przy produkcji nowoczesnych smartphone'ów czy tabletów, dopuszczalna jest pewna wadliwość urządzeń, jeżeli cała seria w dużej mierze jest sprawna. W takich przypadkach wskazana jest kontrola FDP lub FDR.

Procedury jednokrokowe

Z historycznego punktu widzenia, pierwszą korektę poziomu istotności dla zagadnienia testowania zbioru hipotez wprowadził w 1936 r. Bonferroni [3], dzieląc poziom istotności przez liczbę rozpatrywanych hipotez. W ten sposób ustanowił, dla każdej z rozważanych m hipotez, poziom istotności dany wzorem:

$$\alpha_i = \frac{\alpha}{m}.$$

W przypadku procedury Bonferroniego kontroli miary błędu FWER, dopasowane p-wartości:

$$\tilde{p}_i = \min(m \cdot p_i, 1).$$

Procedury takie jak Bonferroniego, które testują wszystkie hipotezy zerowe na tym samym poziomie istotności α_0 , dla każdej z hipotez zerowych z osobna, nazywa się *procedurami jednokrokowymi*. W trakcie zastosowania procedury jednokrokowej, wynik pojedynczego testu nie ma wpływu na postępowanie przy weryfikacji następnej hipotezy. Odrzucane są te hipotezy zerowe, dla których p-wartość p_i jest mniejsza od zadanego poziomu istotności α_0 . Korekta Bonferroniego jest z powodzeniem wykorzystywana do dziś.

Dla procedury Bonferroniego kontroli FWER prawdziwe jest następujące (zobacz [12]):

Twierdzenie 1.1. *Dla zagadnienia testowania zbioru hipotez postaci $H_0^i : \theta \in \Theta, i = 1, \dots, m$ oraz odpowiadających im p-wartości spełniających warunek (1.3) ze Stw 1.1, procedura Bonferroniego, odrzucająca H_0^i , gdy $p_i \leq \alpha/m$, zapewnia mocną kontrolę FWER na poziomie α .*

Dowód. Niech $M_0 := \{i : H_0^i \text{ jest prawdziwa}\}$ jak przy definicji mocnej kontroli. Dodatkowo niech m_0 to liczba prawdziwych hipotez wśród m rozpatrywanych.

$$\begin{aligned} \text{FWER} &= \mathbb{P}_\theta(V \geq 1) \leq \mathbb{E}(V) = \sum_{i \in M_0} (\mathbb{I}(p_i \leq \alpha/m)) \\ &= \sum_{i \in M_0} \mathbb{P}_\theta(p_i \leq \alpha/m) \leq \sum_{i \in M_0} \alpha/m = m_0 \frac{\alpha}{m} \leq \alpha. \end{aligned}$$

Pierwsza nierówność to wynik zastosowanej nierówności Markowa, druga to wykorzystany warunek (1.3). Dowody ogólniejszych twierdzeń kontroli k-FWER można znaleźć w [7, 12]. \square

Procedury wielokrokowe typu step-down

Ponieważ procedura Bonferroniego, dla dużej ilości hipotez, cechuje się małą szansą na odrzucenie hipotezy zerowej, udoskonalono procedury jednokrokowe poprzez nadanie specjalnego schematu postępowania nazwanego *procedurami wielokrokowymi*, które wykorzystują informacje o łącznym rozkładzie p-wartości dla wszystkich hipotez. Procedury wielokrokowe dominują nad procedurami jednokrokowymi, dzięki większym obszarom odrzuceń fałszywych hipotez zerowych. Procedury wielokrokowe wyznaczają kolejność testowania hipotez zerowych według kolejności odpowiadających im p-wartości. Takie postępowanie narzuca ponumerowanie rozważanych hipotez zerowych $H_0^1, H_0^2, \dots, H_0^m$ według odpowiadających im p-wartości numerowanych niemalejąco. Wtedy, dla uporządkowanych p-wartości według reguły:

$$p^{(1:m)} \leq p^{(2:m)} \leq \dots \leq p^{(m:m)},$$

hipoteza $H_0^{(i:m)}$ odpowiada p-wartości $p^{(i:m)}$. Dla procedur wielokrokowych zakłada się, że poziomy istotności tworzą niemalejący ciąg, który można ponumerować tak by zachodziło:

$$0 < \alpha^{(1:m)} \leq \alpha^{(2:m)} \leq \dots \leq \alpha^{(m:m)} < 1.$$

Notacja. Górny indeks $(i:m)$ oznacza i -tą statystykę pozycyjną.

Przykładem procedury wielokrokowej jest procedura typu *step-down* z zadanymi przez Holma w 1979 r. [10] poziomami istotności dla poszczególnych hipotez zerowych, przy założeniu, że statystyki testowe są niezależne, według wzoru:

$$\alpha^{(i:m)} = \frac{\alpha}{m - i + 1}. \quad (1.5)$$

Na cześć autora tę procedurę nazwano **Procedurą Holma** kontroli miary błędu FWER.

Schemat procedury typu step-down

0. Testowanie zaczyna się od $H_0^{(1:m)}$ - hipotezy, dla której otrzymano najmniejszą p-wartość.
1. Jeżeli $p^{(1:m)} > \alpha^{(1:m)}$ to przyjmuje się wszystkie m hipotez.
2. W przeciwnym wypadku sprawdza się kolejno dla $i = 2, \dots, m$ czy $p^{(i:m)} \leq \alpha^{(i:m)}$, aż do momentu uzyskania dla pierwszego r warunku $p^{(r:m)} > \alpha^{(r:m)}$.
3. W tym momencie procedura się kończy i odrzucane są hipotezy $H_0^{(1:m)}, H_0^{(2:m)}, \dots, H_0^{(r-1:m)}$.

Postępowanie według ww. schematu sprawia, że dla procedury Holma kontroli FWER prawdziwe jest następujące twierdzenie (zobacz [12]):

Twierdzenie 1.2. *Dla zagadnienia testowania zbioru hipotez postaci $H_0^i : \theta \in \Theta, i = 1, \dots, m$ oraz odpowiadających im p-wartości spełniających warunek (1.3) ze Stw 1.1, procedura Holma zapewnia mocną kontrolę FWER na poziomie α .*

Dowód. Niech podobnie $M_0 := \{i : H_0^i \text{ jest prawdziwa}\}$ oraz m_0 oznacza liczbę prawdziwych hipotez wśród m rozpatrywanych. Wypada założyć $m_0 \geq 1$, bo inaczej nie ma czego dowodzić. Uporządkujmy p-wartości odpowiadające prawdziwym hipotezom zerowym i wprowadźmy dla nich oznaczenia:

$$q^{(1:m_0)} \leq \dots \leq q^{(m_0:m_0)}.$$

Niech j oznacza najmniejszy indeks, dla którego $p^{(j:m)} = q^{(1:m_0)}$. Wówczas:

$$1 \leq j \leq m - m_0 + 1.$$

Ta nierówność wynika z esencji schematu procedury Holma. Największa możliwa wartość j ($j = m - m_0 + 1$) osiągana jest wtedy, gdy wszystkie najmniejsze p-wartości odpowiadają $m - m_0$ fałszywym hipotezom zerowym, a następne p-wartości odpowiadają m_0 hipotezom prawdziwym. Skoro $p^{(j:m)} = q^{(1:m_0)}$, to procedura Holma kontroli FWER dopuszcza co najwyżej 1 błąd I rodzaju, gdy:

$$p^{(1:m)} \leq \alpha^{(1:m)}, p^{(2:m)} \leq \alpha^{(2:m)}, \dots, p^{(j:m)} \leq \alpha^{(j:m)}.$$

Zatem $q^{(1:m_0)} = p^{(j:m)} \leq \alpha^{(j:m)} = \frac{\alpha}{m-j+1} \leq \frac{\alpha}{m_0}$, bo j maksymalnie $= m - m_0 + 1$.

Ta nierówność oraz rozumowanie podobne jak w Twierdzeniu 1.1 dają:

$$\text{FWER} \leq \mathbb{P}_\theta(q^{(1:m_0)} \leq \frac{\alpha}{m_0}) \leq \alpha.$$

Twierdzenia o kontroli k-FWER dla uogólnionej procedury Holma można znaleźć w [7, 12]. \square

Jak podaje [5], dopasowane p-wartości dla procedury Holma kontroli FWER są określone:

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

Procedury wielokrokowe typu step-up

W 1988 r. Yosef Hochberg [9] zaproponował inny algorytm postępowania procedury wielokrokowej kontroli FWER, przy takich samych poziomach istotności jak w przypadku procedury Holma (1.5) oraz przy założeniu niezależności statystyk testowych. Była to procedura wielokrokowa typu *step-up*, którą można traktować jak odwróconą procedurę typu step-down.

Definicja 1.12. *Procedurę Hochberga kontroli miary błędu FWER na poziomie istotności α typu step-up nazywamy procedurę, która testuje każdą z hipotez odpowiednio na poziomie:*

$$\alpha^{(i:m)} = \frac{\alpha}{m - i + 1}. \quad (1.6)$$

Kluczowym elementem różniącym procedury step-down i step-up jest ich schemat.

Schemat procedury typu step-up

0. Testowanie zaczyna się od $H_0^{(m:m)}$ - hipotezy, dla której otrzymano największą p-wartość.
1. Jeżeli $p^{(m:m)} \leq \alpha^{(m:m)}$ to odrzuca się wszystkie m hipotez.
2. W przeciwnym wypadku sprawdza się kolejno dla $i = m - 1, \dots, 1$ czy $p^{(i:m)} > \alpha^{(i:m)}$, aż do momentu uzyskania dla pierwszego r warunku $p^{(r:m)} \leq \alpha^{(r:m)}$.
3. W tym momencie procedura się kończy i odrzucane są hipotezy $H_0^{(1:m)}, H_0^{(2:m)}, \dots, H_0^{(r:m)}$.

Dla procedury Hochberga kontroli miary FWER prawdziwe jest następujące (zobacz [2]):

Twierdzenie 1.3. *Dla zagadnienia testowania zbioru hipotez postaci $H_0^i : \theta \in \Theta, i = 1, \dots, m$ oraz odpowiadających im p-wartości spełniających warunek (1.3) ze Stw 1.1, procedura Hochberga zapewnia mocną kontrolę FWER na poziomie α .*

Dowód. Niech podobnie $M_0 := \{i : H_0^i \text{ jest prawdziwa}\}$ oraz m_0 oznacza liczbę prawdziwych hipotez wśród m rozpatrywanych. Wypada założyć $m_0 \geq 1$, bo inaczej nie ma czego dowodzić. Uporządkujmy p-wartości odpowiadające prawdziwym hipotezom zerowym i wprowadźmy dla nich oznaczenia:

$$q^{(1:m_0)} \leq \dots \leq q^{(m_0:m_0)}.$$

Niech j oznacza największy indeks, dla którego $p^{(j:m)} = q^{(m_0:m_0)}$. Wówczas:

$$m_0 \leq j \leq m.$$

Najwięcej błędów I rodzaju, bo aż m_0 , jest popełnianych w przypadku, gdy $q^{(m_0:m_0)} \leq \alpha^{(j:m)}$. Wtedy $q^{(m_0:m_0)} = p^{(j:m)} \leq \alpha^{(j:m)} = \frac{\alpha}{m-j+1} \leq \alpha$ bo $m-j+1 \geq 1$ dla $m_0 \leq j \leq m$.

Ta nierówność oraz rozumowanie podobne jak w Twierdzeniach 1.1 i 1.2 dają:

$$\text{FWER} \leq \mathbb{P}_\theta(q^{(m_0:m_0)} \leq \alpha) \leq \alpha.$$

Rozważanie mniejszej niż m_0 liczby popełnionych błędów I rodzaju przez procedurę Hochberga doprowadzi jedynie do otrzymania lepszego ograniczenia kontroli miary błędu w postaci $\frac{\alpha}{2}$ dla $m_0 - 1$ błędów, $\frac{\alpha}{3}$ dla $m_0 - 2$ błędów itp. \square

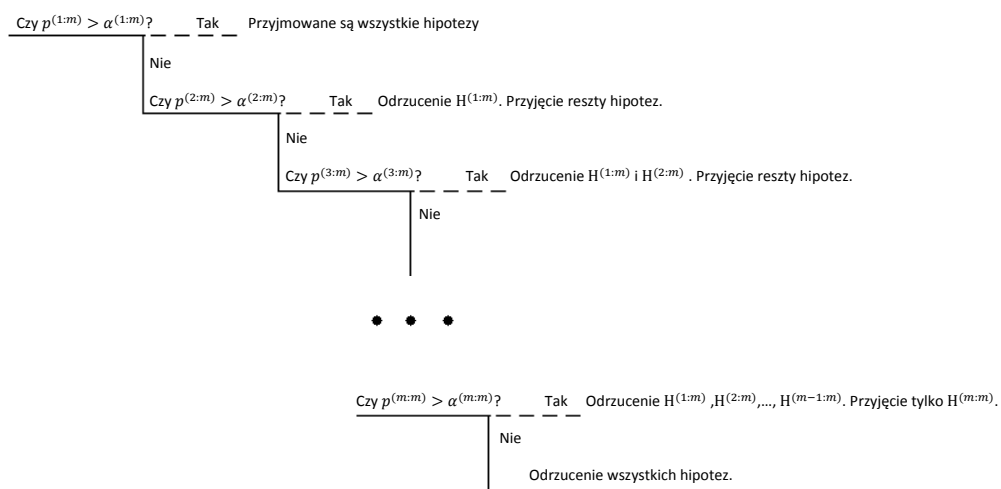
Procedura Hochberga jest praktyczniejszą procedurą testowania niż procedura Holma, gdyż ma większą moc. Oznacza to, że ma większe obszary odrzucenia, przez co odrzuca średnio więcej fałszywych hipotez zerowych.

W przypadku procedury Hochberga kontroli FWER dopasowane p-wartości [5] są zadane:

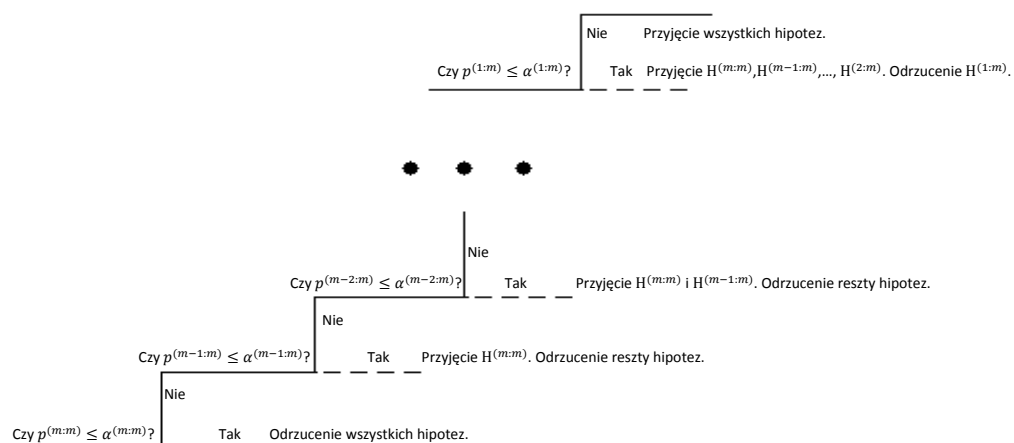
$$\tilde{p}_i = \min_{k=i, \dots, m} \{\min((m-k+1) \cdot p^{(k:m)}, 1)\}.$$

Idee postępowania w przypadku procedur step-up i step-down obrazują Rysunki 1.2 i 1.3.

Więcej informacji na temat rodzajów współczynników miary błędu i procedur ich kontrolujących można znaleźć w [2, 4, 5, 7, 8, 12].



Rysunek 1.2: Schemat procedury typu step-down.



Rysunek 1.3: Schemat procedury typu step-up.

1.3. Test kombinatoryczny Manna-Whitneya-Wilcoxona

W poniższym rozdziale przedstawiono test, którym posłużono się w rozdziale 3 do analizy rzeczywistych danych medycznych. Teoria, dzięki której powstał ten rozdział pochodzi z [18].

Test kombinatoryczny Manna-Whitneya-Wilcoxona (w skrócie MWW) jest testem nieparametrycznym, co oznacza, że nie zakłada znajomości rozkładów badanych próbek. Test MWW określa się mianem testu kombinatorycznego, gdyż zakłada się, że każda permutacja statystyk pozycyjnych z badanej próbki jest jednakowo prawdopodobna. Test MWW weryfikuje hipotezę o parametrze położenia w dwóch próbkach $\mathbb{X} = (X_1, \dots, X_n) \sim F_\mu$ oraz $\mathbb{Y} = (Y_1, \dots, Y_m) \sim F_\nu$, przy czym

$$F_\mu(x) = F(x - \mu) \text{ i } F_\nu(y) = F(y - \nu),$$

dla pewnej (nieznanej) ciągłej dystrybuanty F . Rozpatrywanym zagadnieniem jest hipoteza złożona $H_0^+ : \mu \geq \nu$. Pomysł weryfikacji tej hipotezy jest zawarty w poniższym schemacie testu Manna-Whitneya-Wilcoxona.

Schemat testu Manna-Whitneya-Wilcoxona

1. Należy uporządkować obie próby $\mathbb{X} = (X_1, \dots, X_n)$ i $\mathbb{Y} = (Y_1, \dots, Y_m)$ w jeden ciąg niemalejący (ciągłość dystrybuanty F sprawia z prawdopodobieństwem 1, że ciąg będzie rosnący).
2. Ustala się *rangi* (numery) poszczególnych obserwacji $\mathbb{Y} = (Y_1, \dots, Y_m)$ w tej połączonej próbie nadając Y -om rangi oznaczane S_1, \dots, S_m . Jest jasne, że duże wartości statystyki $W = \sum_{j=1}^m S_j$ świadczą przeciwko H_0^+ .
3. Kolejny krok polega na znalezieniu wartości krytycznej $w_{n,m}(\alpha)$, dla której zachodzi

$$\mathbb{P}_{\mu,\nu}(W > w_{n,m}(\alpha)) \leq \alpha, \quad (1.7)$$

dla wszystkich (μ, ν) takich, że $\mu \geq \nu$. Nie jest to jednak trudne, gdyż

$$\mathbb{P}_F(W > w_{n,m}(\alpha)) = \frac{k(w)}{\binom{m+n}{m}}, \quad (1.8)$$

dla $k(w)$ oznaczającego liczbę kombinacji ustawień \mathbb{X} i \mathbb{Y} w ciąg, dla której $W > w$.

Uzasadnienie. Aby wykazać równość (1.8) zauważmy, że dla każdej liczby $c \geq 0$ zachodzi:

$$W(X_1, \dots, X_n, Y_1, \dots, Y_m) \leq W(X_1, \dots, X_n, Y_1 + c, \dots, Y_m + c).$$

Zatem, przy prawdziwej hipotezie $H_0^+ : \mu \geq \nu$, zachodzi:

$$\mathbb{P}(W(X_1, \dots, X_n, Y_1, \dots, Y_m) > w) \leq \mathbb{P}(W(X_1, \dots, X_n, Y_1 + (\mu - \nu), \dots, Y_m + (\mu - \nu)) > w).$$

Skoro $\mathbb{Y} = (Y_1, \dots, Y_m)$ jest próbą z rozkładu F_ν to $\mathbb{Y}_{\mu-\nu} = (Y_1 + (\mu - \nu), \dots, Y_m + (\mu - \nu))$ pochodzi z rozkładu F_μ , czyli $\mathbb{P}_{\mu,\nu}(W > w) \leq \mathbb{P}_{\mu,\mu}(W > w)$ dla wszystkich (μ, ν) które spełniają $\mu \geq \nu$. Stąd warunek (1.7) można przedstawić w postaci

$$\mathbb{P}_{\mu,\mu}(W > w) \leq \alpha.$$

Jest to prawdopodobieństwo zdarzenia $W > w$ dla dwóch próbek pochodzących z pewnego, tego samego ciągłego rozkładu F , co oznaczyć można w skrócie \mathbb{P}_F . Skoro $\mathbb{X} = (X_1, \dots, X_n)$ oraz $\mathbb{Y} = (Y_1, \dots, Y_m)$ pochodzą z tego samego rozkładu, to wartość statystyki W nie zależy od rozkładu F . Każda kombinacja ustawienia tych dwóch próbek jest jednakowo prawdopodobna, stąd otrzymujemy (1.8).

1.4. Współczynnik korelacji Pearsona

W niniejszym rozdziale opisano narzędzie statystyczne jakim jest współczynnik korelacji liniowej (zwany niekiedy współczynnikiem korelacji Pearsona) oraz związana z nim kowariancja. Współczynnik korelacji posłuży do wyeliminowania pewnych badanych danych, które są zależne od innych. Proces analizy korelacji został przedstawiony w rozdziale 3.2. Poniżej zaprezentowano teoretyczny opis kowariancji i współczynnika korelacji.

Definicja 1.13. *Kowariancją zmiennych losowych X i Y nazywa się liczbę $Cov(X, Y)$ określoną wzorem*

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = \mathbb{E}(XY) - \mathbb{E}X \cdot \mathbb{E}Y.$$

Kowariancję traktuje się jako pewną miarę zgodności dwóch zmiennych losowych z racji tego, że jest średnią wartością iloczynu odchyleń obu zmiennych od ich wartości oczekiwanej. Interpretować kowariancję należy w ten sposób: jeżeli kowariancja jest duża i dodatnia, to zmienne losowe X i Y mają tendencję do przyjmowania równocześnie dużych lub równocześnie małych wartości, zaś jeżeli kowariancja jest duża, co do wartości bezwzględnej, ale ujemna, to zmienna losowa X ma tendencję do przyjmowania wartości dużych, gdy zmienna losowa Y przyjmuje wartości małe i odwrotnie.

Aby wiedzieć, jak duże wartości kowariancji należy zaobserwować, by stwierdzić zależność między zmiennymi losowymi, wprowadzono pojęcie współczynnika korelacji, o którym mówi poniższa definicja.

Definicja 1.14. *Współczynnikiem korelacji X i Y nazywa się liczbę $\rho(X, Y)$ określoną wzorem*

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X) \cdot Var(Y)}}.$$

Współczynnik korelacji mieści się w zbiorze $[-1, 1]$ i nie zależy od przyjętej skali oraz od położenia od początku układu współrzędnych, w którym rejestrowane są zmienne losowe X i Y . Im wartość współczynnika jest, co do modułu, bliższa liczbie 1 tym obie zmienne są mocniej zależne od siebie.

Więcej na temat kowariancji i współczynnika korelacji można znaleźć w [11].

Rozdział 2

Funkcje do testowania zbioru hipotez w pakiecie \mathcal{R}

W poniższym rozdziale przedstawiono funkcje pakietu statystycznego \mathcal{R} ([16]) do przeprowadzania wcześniej omówionego testu Manna-Whitneya-Wilcoxona oraz do testowania zbioru hipotez przy użyciu procedur: Bonferroniego, Holma i Hochberga. Do przeprowadzenia szczegółowej analizy danych w rozdziale 3 posłużą omówione w tym rozdziale dodatki do pakietu \mathcal{R} : pakiet `stats` (sekcja 2.1) oraz pakiet `multtest` (sekcja 2.2). W niniejszym rozdziale nacisk został położony na przedstawienie technicznych aspektów testowania. Źródłem informacji do napisania poniższego rozdziału były [6, 16, 15].

Dla rozróżnienia tekstu od kodu pakietu \mathcal{R} , kod będzie podawany w ramach jak poniżej.

```
przykładowy kod pakietu R
```

2.1. Pakiet `stats`

Pakiet `stats` jest oryginalnie wbudowany do podstawowej wersji pakietu statystycznego \mathcal{R} . Jego twórcami jest zespół tworzący pakiet \mathcal{R} , czyli R Core Team. Pakiet `stats` zawiera funkcje do obliczeń statystycznych i do generowania liczb losowych. Spośród wielu funkcji należących do pakietu przedstawiono w pracy te, które będą wykorzystane do dalszej analizy danych.

2.1.1. Funkcja `wilcox.test()`

Test kombinatoryczny Manna-Whitneya-Wilcoxona przeprowadza się przy użyciu funkcji `wilcox.test`. Funkcja ta służy do przeprowadzenia testu dla jednej próbki oraz dla dwóch próbek. W tej pracy skupiono się na testach dwópróbkowych. Wywołanie funkcji `wilcox.test` w pakiecie \mathcal{R} wygląda następująco:

```
wilcox.test(x, y,
            alternative = c("two.sided", "less", "greater"),
            mu = 0, paired = FALSE, exact = NULL, correct = TRUE, ...)
```

Argumenty funkcji `wilcox.test` są objaśnione w Tabeli 2.1.

Argument	Opis
<code>x</code>	wektor obserwacji jednowymiarowych z pierwszej próbki
<code>y</code>	wektor obserwacji jednowymiarowych z drugiej próbki
<code>alternative</code>	określenie hipotezy alternatywnej spośród <code>"two.sided"</code> (domyślnie), <code>"less"</code> i <code>"greater"</code>
<code>paired</code>	wartość logiczna precyzująca, czy test ma być sparowany
<code>exact</code>	wartość logiczna, mówiąca czy podawać p-wartość (domyślnie <code>FALSE</code>)
<code>mu</code>	liczba określającą zakładaną różnicę między parametrami położenia <code>x</code> i <code>y</code> (domyślnie 0)
<code>correct</code>	wartość logiczna precyzująca, czy zastosować ciągłą korektę w normalnej aproksymacji p-wartości

Tabela 2.1: Opis argumentów funkcji `wilcox.test`.

Funkcja `wilcox.test` pomija puste/brakujące dane dla argumentów `x` i `y`. Dla wywołania tej funkcji przy podawaniu argumentu `alternative` wystarczy podać pierwszą literę nazwy dla rozważanej hipotezy alternatywnej, czyli zamiast `"greater"` – `"g"` i odpowiednio zamiast `"less"` – `"l"`. Gdy argument `paired` jest ustawiony na wartość logiczną `TRUE` to hipoteza zerowa zakłada, że zakładana w teście dystrybuanta dla `"x-y"` jest symetryczna wokół `mu`. W przeciwnym wypadku, gdy dla argumentu `paired` podana jest wartość logiczna `FALSE` to przeprowadzany jest typ test Manna-Whitneya-Wilcoxona omawiany w podrozdziale 1.3. Dla tego przypadku, hipoteza zerowa zakłada, że dystrybuanta `x` różni się od dystrybuanty `y` o pewną lokację zadaną przez `mu`. Jeżeli próbki są mniejsze niż 50 obserwacji, to jest podawana p-wartość, gdy argument `exact` jest pominięty lub ustawiony na wartość logiczną `TRUE`.

2.1.2. Obiekt klasy `"htest"`

Funkcja `wilcox.test` jest obiektem klasy `"htest"`. Oznacza to, że ta funkcja zwraca listę wartości, do których można się odwołać. Opis wybranych wartości obiektu klasy `"htest"` jest przedstawiony w Tabeli 2.2.

Wartość	Opis
<code>statistic</code>	wartość statystyki testowej
<code>p.value</code>	p-wartość testu
<code>null.value</code>	wartość parametru <code>mu</code>
<code>alternative</code>	obiekt typu <code>string</code> określający hipotezę alternatywną
<code>method</code>	obiekt typu <code>string</code> mówiący jaki typ testu został przeprowadzony
<code>data.name</code>	obiekt typu <code>string</code> podający nazwy danych wykorzystanych w teście

Tabela 2.2: Opis wybranych wartości obiektu klasy `"htest"`.

Odwołania do wartości można dokonać poprzez dodanie do kodu funkcji znaku `$` oraz wskazaniu nazwy żądanej wartości. Poniżej przykład odwołania do p-wartości testu.

```
wilcox.test(argumenty funkcji)$p.value
```

2.1.3. Funkcja cor() i funkcja pairs()

Do zbadania współczynnika korelacji Pearsona między zmiennymi wykorzystanymi w analizie użyto w rozdziale 3 funkcji `cor`. Kod generujący tę funkcję wygląda następująco:

```
cor(x, y = NULL, use = "everything",  
    method = c("pearson", "kendall", "spearman"))
```

Lisa argumentów funkcji `cor` wraz z objaśnieniami została ukazana w Tabeli 2.3.

Argument	Opis
<code>x</code>	wektor lub macierz klasy <code>numeric</code> oznaczająca zbiór obserwacji z pierwszej próbki
<code>y</code>	wektor lub macierz klasy <code>numeric</code> oznaczająca wektor obserwacji z drugiej próbki, równych wymiarów co <code>x</code>
<code>use</code>	określenie metody obliczania korelacji w przypadku brakujących danych, wykorzystano w pracy "pairwise.complete.obs"
<code>method</code>	określenie rodzaju wyliczanego współczynnika korelacji spośród "pearson" (domyślnie), "kendall", lub "spearman"

Tabela 2.3: Opis argumentów funkcji `cor`.

W dalszej części pracy współczynnik korelacji w przypadku brakujących danych będzie liczony tylko dla niewybrakowanych par obserwacji (`use = "pairwise.complete.obs"`). Funkcja `cor` zostanie użyta w funkcji `panel.pearson`, która zostanie wykorzystana jako argument `upper.panel` w funkcji `pairs` do graficznego przedstawienia rozrzutu sparowanych zmiennych oraz ich współczynnika korelacji. Kod funkcji prezentuje się następująco:

```
panel.pearson <- function(x, y, ...) {  
  horizontal <- (par("usr")[1] + par("usr")[2]) / 2;  
  vertical <- (par("usr")[3] + par("usr")[4]) / 2;  
  text(horizontal, vertical,  
        format(abs(cor(x,y,use = "pairwise.complete.obs")), digits=2)) }  
}
```

Argumenty `x` i `y` mają to samo znaczenie co w Tabeli 2.3. Z kolei wywołanie funkcji `pairs` z pakietu graficznego `graphics` wygląda tak:

```
pairs(x, labels, panel = points, ..., lower.panel = panel, upper.panel = panel,  
      diag.panel = NULL, text.panel = textPanel, font.labels = 1, gap = 1)
```

Opis potrzebnych argumentów funkcji `pairs` zawarty został w Tabeli 2.4.

Argument	Opis
<code>x</code>	nazwy punktów, których współrzędne zostaną przedstawione w postaci wykresów rozrzutu, podane w wg formuły $\sim x + y + z$
<code>data</code>	ramka danych, z której są brane zmienne do przedstawienia
<code>labels</code>	określenie nazw zmiennych wyświetlanych na wykresie
<code>panel</code>	określenie funkcji wykorzystanej do przedstawienia informacji na panelach, panele <code>:lower</code> - pod przekątną, <code>upper</code> - nad przekątną, <code>diag</code> na przekątnej
<code>gap</code>	dystans pomiędzy podwykresami
<code>font.labels</code>	rodzaj czcionki podpisów na przekątnej powstałego wykresu

Tabela 2.4: Opis argumentów funkcji `pairs`.

2.2. Pakiet multtest

Pakiet `multtest` zawiera kolekcję funkcji do testowania zbioru hipotez. Zostały w nim zaimplementowane, między innymi, procedury kontroli miary błędu FWER, który został omówiony w podrozdziale 1.2.3. Interesujące z punktu widzenia tej pracy procedury zawarte w pakiecie `multtest` to procedury: Bonferroniego, Hochberga(1988) oraz Holma(1979). Wyniki procedur są przedstawiane za pomocą dopasowanej p-wartości omówionej w podrozdziale 1.2.6. Pakiet ten jest rozszerzeniem pakietu `R` i wymaga dodatkowej instalacji, jednak nie jest obecnie dostępny do pobrania z serwerów CRAN. Twórcy pakietu Katherine S. Pollard, Houston N. Gilbert, Yongchao Ge, Sandra Taylor oraz Sandrine Dudoit udostępniają ten pakiet do pobrania na swojej stronie z oprogramowaniem bioinformatycznym [15]. Z wielu funkcji zawartych w pakiecie `multtest` omówiono tylko te, które wykorzystane zostaną do analizy danych zawartej w rozdziale 3. Pakiet `multtest` jest bardzo funkcjonalny – funkcje, które przedstawiono to tylko jego niewielka część. Więcej informacji można uzyskać w [6].

2.2.1. Funkcja `mt.rawp2adjp()`

Pakiet `multtest` zawiera bardzo przydatną funkcję `mt.rawp2adjp`, dzięki której można przekształcić wejściowy wektor nieprzetworzonych p-wartości na wektor dopasowanych p-wartości. Stąd nazwa – ang. *raw* oznacza nieprzetworzone/surowe, zaś *adj* to skrót od ang. *adjusted* czyli dopasowane. Tej funkcji można użyć, między innymi, dla procedur Bonferroniego, Hochberga i Holma kontroli FWER. Wywołanie tej funkcji wygląda następująco:

```
mt.rawp2adjp(rawp, proc=c("Bonferroni", "Holm", "Hochberg",  
                          "SidakSS", "SidakSD", "BH", "BY", "ABH", "TSBH"),  
             na.rm = FALSE)
```

Objaśnienia argumentów są przedstawione w Tabeli 2.5.

Argument	Opis
<code>rawp</code>	wektor klasy <code>numeric</code> zawierający nieprzetworzone p-wartości
<code>proc</code>	wektor klasy <code>string</code> zawierający nazwy procedur do przeprowadzenia
<code>na.rm</code>	wartość logiczna precyzująca czy pomijać brakujące dane w wektorze <code>rawp</code>

Tabela 2.5: Opis argumentów funkcji `mt.rawp2adjp`.

Dla argumentu `proc` można podać więcej nazw procedur do testowania zbioru hipotez niż wykorzystywane w tej pracy później procedury Bonferroniego, Hochberga czy Holma. Wyprecyzowanie wszystkich możliwych typów procedur dla argumentu `proc` jest celowym zabiegiem mającym ukazać ogrom możliwości jakie daje funkcja `mt.rawp2adjp`. Funkcja `mt.rawp2adjp` zwraca listę wartości, które można wykorzystać. Dwie ciekawe wartości to `adjp` i `index`. Wartość `adjp` zwraca macierz dopasowanych p-wartości, której rzędy odpowiadają hipotezom, a kolumny procedurom testowania zbioru hipotez. Hipotezy są posortowane rosnąco według ich początkowych p-wartości. Z kolei wartość `index` to wektor wskaźników do wierszy wektora `rawp`, o długości pomiędzy 1 a długością wektora `rawp`, którego wiersze są posortowane według wejściowych p-wartości. Aby uzyskać dopasowane p-wartości w kolejności oryginalnych danych należy użyć polecenia `adjp[order(index),]`.

2.2.2. Funkcja `mt.reject()`

Funkcja `mt.reject` zwraca wartości logiczne odzwierciedlające czy odrzucono daną procedurę i liczbę odrzuconych hipotez dla kilku procedur testowania zbioru hipotez i różnych poziomów kontroli miary FWER.

```
mt.reject(adjp, alpha)
```

Wyjaśnienie argumentów tej funkcji jest przedstawione w Tabeli 2.6.

Argument	Opis
<code>adjp</code>	macierz dopasowanych p-wartości uzyskana przy użyciu <code>mt.rawp2adjp</code>
<code>alpha</code>	liczba lub wektor liczb określający poziomy kontroli miary błędu FWER

Tabela 2.6: Opis argumentów funkcji `mt.reject`.

Funkcja `mt.reject` zwraca dwie wartości `r` i `which`, które są macierzami. Macierz `r` zawiera liczby odrzuconych hipotez dla kilku procedur testowania zbioru hipotez i różnych poziomów kontroli miary FWER. Wiersze `r` odpowiadają poziomom kontroli, a kolumny procedurom testowania zbioru hipotez. Natomiast macierz `which` to macierz wartości logicznych, oznaczających czy dokonano odrzucenia poszczególnych hipotez, dla różnych procedur testowania zbioru hipotez, przy poziomie kontroli `alpha[1]` – poziom kontroli stojący na pierwszym miejscu wektora `alpha` lub w przypadku podania argumentu `alpha` jako liczby, to dokładnie liczba `alpha`. Wiersze odpowiadają hipotezom, a kolumny procedurom testowania zbioru hipotez.

2.2.3. Funkcja `mt.plot()`

Pakiet `multtest` udostępnia funkcję graficzną do przedstawiania, wyliczonych za pomocą funkcji `mt.rawp2adjp` i `mt.reject`, danych na wykresie. Wywołanie tej funkcji przedstawiano poniżej:

```
mt.plot(adjp, teststat, plottype="rvsa", logscale=FALSE,  
        alpha=seq(0, 1, length = 100), proc, leg=c(0, 0), ...)
```

Argumenty opisano i przedstawiono w Tabeli 2.7.

Argument	Opis
<code>adjp</code>	macierz dopasowanych p-wartości uzyskana przy użyciu <code>mt.rawp2adjp</code>
<code>teststat</code>	wektor statystyk testowych każdej z hipotez, można uzyskać przy pomocy funkcji <code>mt.teststat</code> lub <code>wilcox.test</code>
<code>plottype</code>	określenie typu wykresu spośród 4 dostępnych: <code>rvsa</code> , <code>pvsr</code> , <code>pvst</code> i <code>pvsj</code>
<code>logscale</code>	wartość logiczna określająca czy zastosować skalę przeciwnie logarytmiczną w wykresach typu <code>pvsj</code> i <code>pvst</code>
<code>alpha</code>	wektor wartości współczynników kontroli FWER dla typu <code>rvsa</code>
<code>proc</code>	procedury do podania w legendzie
<code>leg</code>	współrzędne legendy
<code>...</code>	własności wykresu typu podpis, grubość i kolor linii

Tabela 2.7: Opis argumentów funkcji `mt.plot`.

Argument `plotttype` ustawiony na `rvsa` generuje wykres zależności odrzuconych hipotez zerowych od ustalonego poziomu miary błędu FWER. Ustawienie na `pvsr` powoduje powstanie wykresu zależności miary błędu FWER od liczby odrzuconych hipotez, przy wyrysowaniu dopasowanych p-wartości. Z kolei ustawienie na `pvsr` umożliwia stworzenie wykresu obrazującego dopasowane p-wartości dla każdej z podanych procedur kontroli, a ustawienie na `pvst` generuje wykres zależności dopasowanych p-wartości od statystyk testowych.

Ponieważ funkcja `mt.plot` ma z góry ustalone podpisy osi na wykresie w języku angielskim i nie można ich zmienić, w pracy, aby uzyskać wykresy z polskimi podpisami, wykorzystano bliźniaczą, specjalnie stworzoną na te potrzeby funkcję `mt.plot2`, która posiada takie same argumenty jak funkcja `mt.plot` z tą różnicą, iż udostępnia możliwość dodania własnych podpisów osi. Zastosowano jeszcze jedną drobną ingerencję w kodzie funkcji `mt.plot`. Po przypisaniu jej do nowej funkcji o nazwie `mt.plot2` niemożliwe było jej pełne wywołanie ze względu na zawartą w niej, niedostępną dla użytkowników, funkcję `mt.legend`. Powodowało to brak rysowania legendy na wykresie. Zastosowano przymusowe przypisanie funkcji `mt.legend` do nowej funkcji `legenda` i zamieniono ze sobą te funkcje wewnątrz kodu funkcji `mt.plot2`. Szczegóły można znaleźć w Dodatku A.

2.2.4. Porównanie procedur Hochberga i Holma przy użyciu pakietu `multtest`

Przy pomocy funkcji `mt.rawp2adjp` i `mt.reject` można dokonać porównania procedur Hochberga i Holma kontroli błędu FWER. Dla zadanego na potrzeby tego przykładu wektora p-wartości `przykladP` danego jak poniżej¹

```
> przykladP <- c(0.018, 0.01, 0.023, 0.11)
```

wyliczono dopasowane p-wartości dla poziomu kontroli miary błędu FWER $\alpha = 0.05$

```
> mt.rawp2adjp(przykladP, proc=c("Holm", "Hochberg"))
$adjp
      rawp  Holm Hochberg
[1,] 0.010 0.040    0.040
[2,] 0.018 0.054    0.046
[3,] 0.023 0.054    0.046
[4,] 0.110 0.110    0.110

$index
[1] 2 1 3 4
```

oraz użyto funkcji `mt.reject` do sprawdzenia prawdziwości przykładowych hipotez.

```
> mt.reject(mt.rawp2adjp(przykladP, proc=c("Holm", "Hochberg"))$adjp,
+           alpha=c(0.05, 0.1))
$r
      rawp  Holm Hochberg
0.05     3     1         3
0.1      3     3         3

$which
      rawp  Holm Hochberg
[1,] TRUE  TRUE    TRUE
[2,] TRUE FALSE    TRUE
[3,] TRUE FALSE    TRUE
[4,] FALSE FALSE   FALSE
```

¹Wprowadzane komendy do \mathcal{R} oznaczane są na początku linii znakiem ">". Linie kodu bez tego znaku to odpowiedź wygenerowana przez pakiet \mathcal{R} .

W powyższym przykładzie wartość TRUE dla macierzy `which`, która jest wynikową wartością funkcji `mt.reject` oznacza, że daną hipotezę odrzucono. Jak widać procedura Hochberga ma większą moc, gdyż odrzuciła więcej hipotez (pierwsze 3) niż procedura Holma, która odrzuciła tylko pierwszą. Dopasowane p-wartości były porównywane z zadaniem poziomem kontroli miary błędu FWER $\alpha = 0.05$. Odrzucenie hipotezy następowało w momencie, gdy dana dopasowana p-wartość była mniejsza od $\alpha = 0.05$.

Wartość `r` to macierz oznaczająca ile hipotez odrzucono dla każdej z procedur na każdym z rozważanych dwóch poziomów kontroli $\alpha = 0.05$ i $\alpha = 0.10$. Wartość `index` zwraca kolejność p-wartości przed uporządkowaniem.

Dla wejściowych, niezmodyfikowanych p-wartości należałoby porównać p-wartości

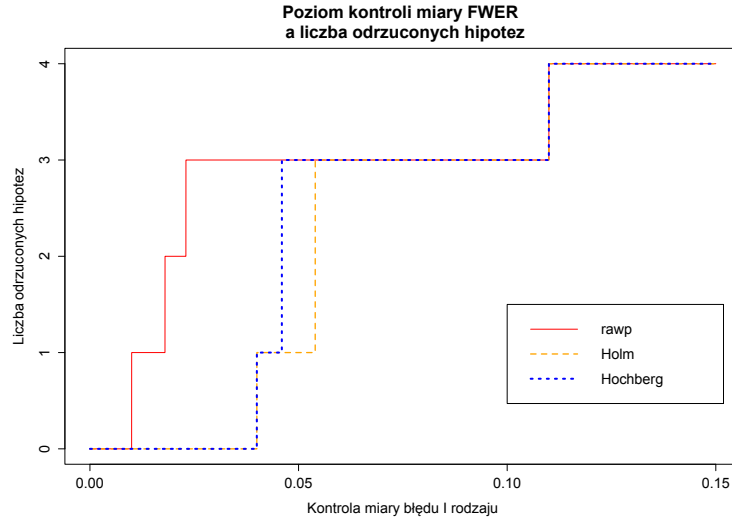
$$p^{(1:4)} = 0.01, p^{(2:4)} = 0.018, p^{(3:4)} = 0.023, p^{(4:4)} = 0.11$$

z odpowiednio wyliczonymi wg reguły $\alpha^{(i:4)} = \alpha/(4-i+1)$ poziomami istotności. Wtedy, dla

$$\alpha^{(1:4)} = 0.0125, \alpha^{(2:4)} = 0.0167, \alpha^{(3:4)} = 0.025, \alpha^{(4:4)} = 0.05$$

procedura Holma, jako procedura step-down odrzuca tylko H_1 , gdyż $p^{(1:4)} \leq \alpha^{(1:4)}$ i zatrzymuje się bowiem $p^{(2:4)} > \alpha^{(2:4)}$. Z kolei procedura Hochberga, jako procedura typu step-up przyjmuje H_4 , gdyż $p^{(4:4)} > \alpha^{(4:4)}$ i zatrzymuje się odrzucając wszystkie inne procedury ponieważ $p^{(3:4)} \leq \alpha^{(3:4)}$.

Dla tego przykładu, przy pomocy funkcji `mt.plot2`, można stworzyć wykres przedstawiający zależność liczby odrzuconych hipotez zerowych od zadanego poziomu miary błędu FWER - Rysunek 2.1. Kod jest dostępny w Dodatku A.



Rysunek 2.1: Poziom kontroli miary FWER a liczba odrzuconych hipotez dla procedur Holma (na żółto) i Hochberga (na niebiesko) - przykład. Czerwonym kolorem ukazano odrzucenia dla niezastosowanej poprawki poziomów istotności pojedynczych testów w przypadku zagadnienia testowania zbioru hipotez.

Rozdział 3

Analiza danych medycznych

Niniejszy rozdział poświęcony jest analizie danych. W sekcji 3.1 przedstawiono opis analizowanych danych. W sekcji 3.2 zbadano korelacje między badanymi zmiennymi. Sekcja 3.3 prezentuje wyniki testów Manna-Whitneya-Wilcozona porównujących parametry serca, zaś sekcja 3.4 poświęcona jest zagadnieniu testowania zbioru hipotez dla wyników z sekcji 3.3.

3.1. Opis danych

Dane wykorzystane w pracy, na których przedstawiono zagadnienie testowania zbioru hipotez procedurami Bonferroniego, Hochberga i Holma kontroli FWER, zostały zebrane przez **Oddział Kardiologii Inwazyjnej** Warszawskiego Uniwersytetu Medycznego i dotyczą dwóch metod przeprowadzania zabiegów w przypadku wystąpienia zagrożenia życia pacjenta jakim jest **zawał**. Celem pracy jest porównanie skuteczności dwóch metod zabiegów, których dotyczą dane:

metody "balonikowania"	fachowo nazywanej <i>Przezskórną angioplastyką wieńcową</i>
metody "bypassów"	fachowo nazywanej <i>Pomostowaniem aortalno-wieńcowym</i> .

Dane przedstawiają wybrane parametry serca zmierzone przed przeprowadzeniem zabiegu nazwane: EF, AVA, Max i Mean oraz wybrane parametry serca zmierzone po przeprowadzeniu zabiegu nazwane: AVA.post, Max.post i Mean.post. Dane składają się z 67 obserwacji i 9 zmiennych (7 wymienionych parametrów serca, wiek oraz typ zabiegu). Każda obserwacja odpowiada jednemu pacjentowi. W pracy nie analizowano wieku pacjentów, ze względu na małą próbkę danych do badania jakichkolwiek zależności od wieku. Wejściowy zbiór danych podzielono na 2 podzbiory:

ba1	pacjenci, na których przeprowadzono zabieg balonikowania
nba1	pacjenci, na których przeprowadzono zabieg bypassów

Oznaczenia zmiennych.

EF	oznacza <i>frakcję wyrzutową serca</i> .
AVA i AVA.post ¹	oznacza <i>powierzchnię przekroju ujścia aorty</i> .
Max i Max.post ¹	oznacza <i>maksymalny zmierzony gradient ciśnień</i> .
Mean i Mean.post ¹	oznacza <i>średni zmierzony gradient ciśnień</i> .

¹Zmienne z rozszerzeniem ".post" oznaczają parametry zmierzone po zabiegu, zaś zmienne bez tego rozszerzenia to parametry zmierzone przed zabiegiem.

W krajach europejskich choroby układu krążenia (w szczególności serca, tętnic i żył) są najczęstszą przyczyną zgonów. Z powodu schorzeń sercowo-naczyniowych, w tym zawału mięśnia sercowego, umiera dwa razy więcej osób niż z powodu chorób nowotworowych [13].

Przez zawał rozumie się:

Definicja 3.1. *Zawał mięśnia sercowego, nazywany też zawałem serca i atakiem serca - martwica mięśnia sercowego spowodowana jego niedokrwieniem na skutek zamknięcia tętnicy wieńcowej doprowadzającej krew do obszaru serca.*

Jak podaje [17], w przypadku chorób niedokrwienia serca stosowane leczenie inwazyjne polega, między innymi, na zastosowaniu przezskórnych interwencji wieńcowych. Dwoma zabiegami przezskórnej interwencji wieńcowej są **przezskórna angioplastyka wieńcowa** oraz **pomostowanie aortalno-wieńcowe**. Dane zawarte w pracy dotyczą obu tych zabiegów.

Pomostowanie aortalno-wieńcowe

Klasyczny zabieg polega na wykonaniu, za pomocą sternotomii pośrodkowej, cięcia na sercu o długości 20-25 cm. Po pobraniu żyły odpiszczelowej lub pozyskaniu pomostu tętniczego (najczęściej tętnicy piersiowej wewnętrznej) wprowadza się kaniulę do aorty wstępującej i do prawego przedsionka serca i podłącza krążenie pozaustrojowe. Po wykonaniu zespolenia i ogrzaniu chorego odkleszcza się aortę i przywraca krążenie wieńcowe. Zespolenia pomostów z aortą wstępującą wykonuje się na bijącym sercu, na stycznie zakleszczonej aorcie [17].

Przezskórna angioplastyka wieńcowa

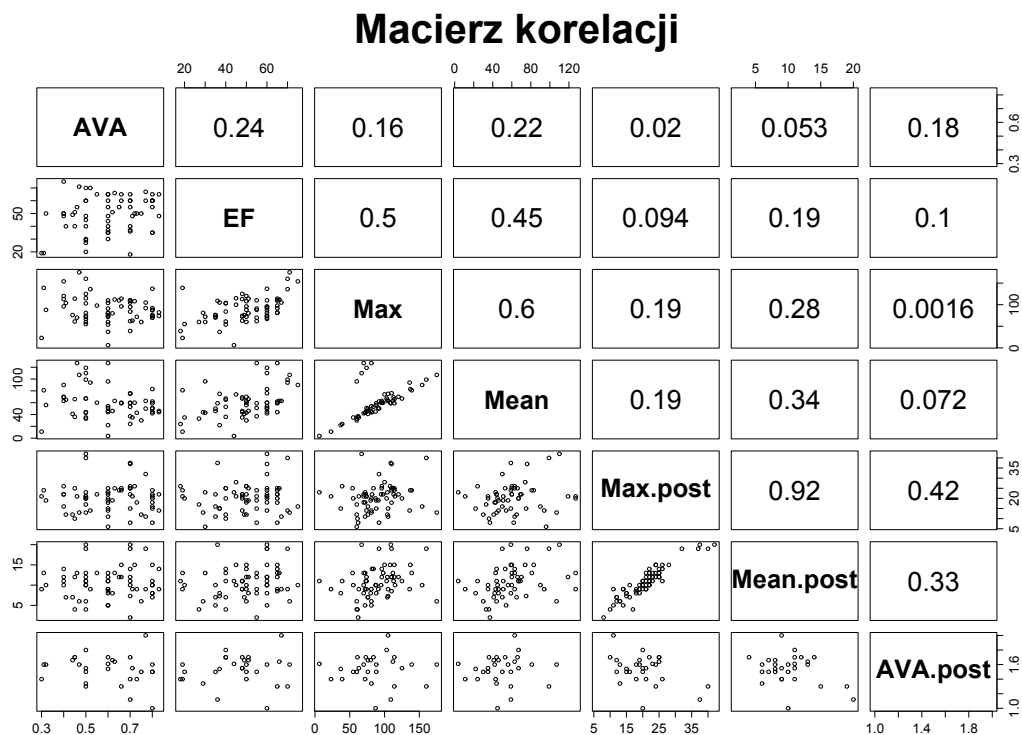
Zabieg polega na poszerzeniu zwężonej tętnicy wieńcowej za pomocą balonu na cewniku wprowadzonym przezskórnie do tętnicy wieńcowej, w celu przywrócenia prawidłowego ukrwienia mięśnia sercowego. Przez nakłucie tętnicy obwodowej, najczęściej tętnicy udowej, wprowadza się cewnik do ujścia zwężonej tętnicy wieńcowej. Przez kanał cewnika wprowadza się prowadnik, przesuwając go pod kontrolą fluoroskopową poza zwężenie tętnicy wieńcowej. Na prowadniku wprowadza się w miejsce zwężenia cewnik z balonem, którego rozmiar jest dobierany do średnicy odcinka tętnicy przed i za zwężeniem. Po dokładnym umiejscowieniu balonu wypełnia się go na kilkadziesiąt sekund płynem pod ciśnieniem od kilku do kilkunastu atmosfer. Jeżeli pierwotny efekt jest niezadowalający, zabieg można powtórzyć, wypełniając balon pod coraz większym ciśnieniem. Optymalny efekt uzyskuje się zwykle po wszczepieniu stentu w miejscu zwężenia [17].

3.2. Analiza korelacji

W tym rozdziale zostanie dokonana analiza korelacji. Zostaną zbadane zależności między zmiennymi opisującymi rozważane w pracy parametry serca. Należy wyeliminować z analizy zmienne, które wykazują dość znaczącą korelację z innymi parametrami, aby móc w dalszej części pracy stosować procedury testowania zbioru hipotez kontroli miary błędu FWER, omówione w podrozdziale 1.2.6, które wymagają, aby statystyki testowe rozważanych hipotez były niezależne. Dzięki funkcji `pairs`, do której zastosowano funkcję `cor`, wyznaczającą współczynnik korelacji Pearsona, uzyskano macierz korelacji rozważanych 7 parametrów serca oraz wykresy rozrzutu między ich wszystkimi parami. Kod generujący funkcję wygląda następująco:

```
pairs(~AVA+EF+Max+Mean+Max.post+Mean.post+AVA.post,
      data=dane.cor, upper.panel=panel.pearson, pch=21, cex=0.8, font.labels=2,
      cex.labels=1.9, main="Macierz_korelacji", cex.main=2)
```

Efekt zastosowania funkcji `pairs` z uwzględnionym argumentem `upper.panel=panel.pearson` został zaprezentowany na Rysunku 3.1.



Rysunek 3.1: Macierz korelacji i wykresy rozrzutu analizowanych parametrów serca. Można stwierdzić wysoką zależność między parametrami Max i Mean oraz między Max.post i Mean.post, z racji na wysokie współczynniki korelacji - 0.60 i 0.92.

Dzięki powyższym wywołaniom funkcji `pairs` wiadomo, że parametr Max jest skorelowany z parametrem Mean oraz parametr Max.post jest silnie skorelowany z parametrem Mean.post. W dalszej części pracy do analizy porównawczej dwóch metod zabiegu nie będą wykorzystywane parametry Mean oraz Mean.post. Zapewni to spełnienie warunku niezależności statystyk testowych wymaganego w procedurach: Bonferroniego, Holma i Hochberga.

3.3. Test kombinatoryczny Manna-Whitneya-Wilcoxona

Do przeprowadzenia analizy porównawczej dwóch metod zabiegów leczenia pacjentów dotkniętych zawałem posłuży test Manna-Whitneya-Wilcoxona, którego teoretyczny zapis jest podany w podrozdziale 1.3. Do przeprowadzenia testu wykorzystana zostanie, omówiona w podrozdziale 2.1.1, funkcja `wilcox.test`. Nieskorelowane parami parametry serca: EF, AVA, Max, AVA.post i Max.post będą porównywane w dwóch grupach pacjentów `bal` i `nbal`. Przy oznaczeniach $(EF, AVA, Max, AVA.post, Max.post) = (1, 2, 3, 4, 5)$, dla każdego porównania danego parametru zakładane będą hipotezy zerowe i alternatywne jak poniżej:

H_0^i : pacjenci z grup `bal` i `nbal` mają średnio **równy** i -ty parametr serca.

H_1^i : pacjenci z grupy `bal` mają średnio **niższy** i -ty parametr serca niż pacjenci z grupy `nbal`.

Oznacza to, że analizowane będą hipotezy zerowe o równości parametru położenia, przeciwko hipotezie alternatywnej o mniejszym parametrze położenia dla pacjentów z grupy `bal`. Tak postawione hipotezy są bezpieczniejszym rozwiązaniem, gdyż lepiej przyjąć hipotezę o równej sprawności metod, niż przyjąć alternatywną mówiącą o mniejszej skuteczności balonikowania, gdyż balonikowanie ma tę zaletę, że jest mniej inwazyjne i może być powtórzone wielokrotnie w krótkim czasie. Wywołanie funkcji `wilcox.test` przeprowadzającej test dla parametru AVA i wyświetlona przez pakiet `R` odpowiedź wygląda następująco:

```
> wilcox.test(bal$AVA, nbal$AVA, alternative="less", paired=FALSE,
+             correct=FALSE, exact=TRUE)

Wilcoxon rank sum test

data:  bal$AVA and nbal$AVA
W = 490.5, p-value = 0.4784
alternative hypothesis: true location shift is less than 0
```

Statystyka testowa testu wyniosła: $W = 490.5$, a p -wartość testu: $p\text{-value} = 0.4784$. Wyniki testów dla pozostałych parametrów zostały przedstawione w poniższej Tabeli 3.1.

	EF	AVA	Max	AVA.post	Max.post
Wartość statystyki	457.5	490.5	409	55.5	475
p-wartość	0.308	0.478	0.100	0.019	0.451

Tabela 3.1: Wyniki przeprowadzonych testów Manna-Whitneya-Wilcoxona dla dwóch próbek: `bal` i `nbal` dla parametrów EF, AVA, Max, AVA.post i Max.post dla hipotez alternatywnych o mniejszym parametrze położenia w grupie `bal`.

Wnioski. Obliczone p -wartości świadczą, że na poziomie istotności $\alpha = 0.05$ nie ma podstaw do odrzucenia hipotez $H_0^1, H_0^2, H_0^3, H_0^5$ dla parametrów EF, AVA, Max oraz Max.post. Wartość krytyczna testu dla parametru AVA.post równa 0.019 daje podstawy do odrzucenia hipotezy zerowej i przyjęcia hipotezy alternatywnej H_1^4 : pacjenci z grupy `bal` mają średnio **niższy** parametr serca AVA.post niż pacjenci z grupy `nbal`. Potwierdzeniem wyniku 5 powyższych testów jest Rysunek 3.2, który przedstawia wykresy pudełkowe badanych parametrów serca: EF, AVA, Max, AVA.post i Max.post odpowiednio w porównaniu dla dwóch grup badanych pacjentów. Dodatkowe poparcie dla wniosków przedstawia Tabela 3.2 prezentująca statystyki agregujące uzyskane przy pomocy funkcji `summary`, której wywołanie i wynik dla grupy `bal` dla parametrów EF, AVA i Max wyglądają tak:

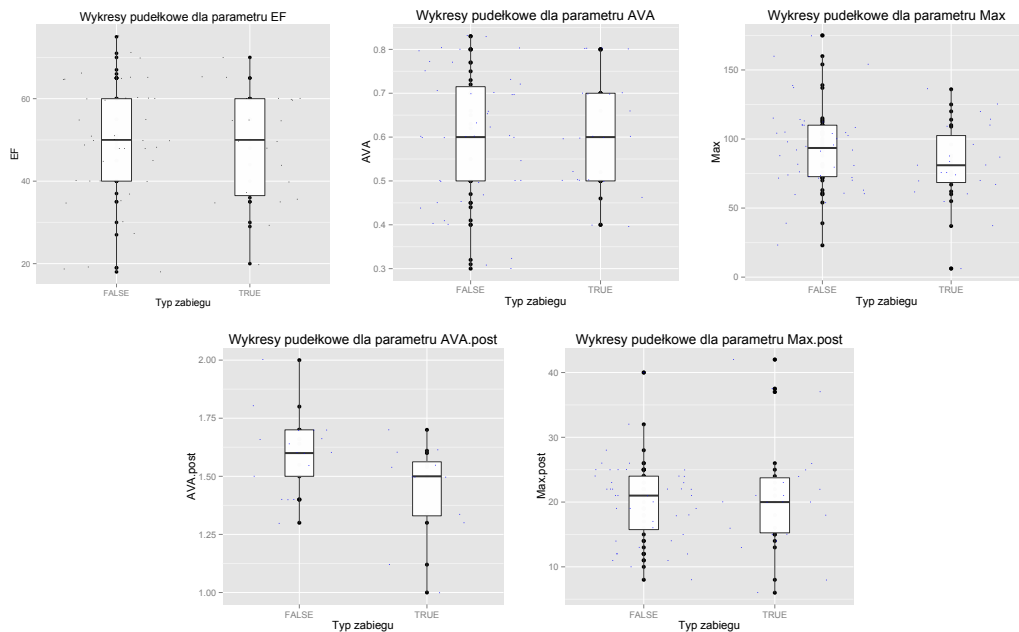
```
> summary(bal[,c(2,3,4)])
```

	EF	AVA	Max
Min.	:20.00	Min. :0.4000	Min. : 6.20
1st Qu.:	36.50	1st Qu.:0.5000	1st Qu.: 68.50
Median :	50.00	Median :0.6000	Median : 81.00
Mean :	48.65	Mean :0.6061	Mean : 81.92
3rd Qu.:	60.00	3rd Qu.:0.7000	3rd Qu.:102.50
Max.	:70.00	Max. :0.8000	Max. :136.00

Parametr	EF		AVA		Max		AVA.post		Max.post	
Typ Zabiegu	nbal	bal	nbal	bal	nbal	bal	nbal	bal	nbal	bal
Minimum	18.00	20.00	0.30	0.40	23.00	6.20	1.30	1.00	8.00	6.00
Q_1	40.00	36.50	0.50	0.50	72.75	68.50	1.50	1.33	15.75	15.25
Mediana	50.00	50.00	0.60	0.60	93.50	81.00	1.60	1.50	21.00	20.00
Średnia	50.16	48.65	0.60	0.60	93.55	81.92	1.60	1.43	20.20	21.07
Q_3	60.00	60.00	0.71	0.70	110.00	102.50	1.70	1.56	24.00	23.75
Maksimum	75.00	70.00	0.83	0.80	175.00	136.00	2.00	1.70	40.00	42.00

Tabela 3.2: Statystyki agregujące parametrów serca: EF, AVA, Max, AVA.post i Max.post w dwóch grupach pacjentów. Q_1 oznacza pierwszy kwartył, Q_3 oznacza trzeci kwartył.

Wszystkie statystyki agregujące parametru Max w grupie pacjentów **bal** są niższe od odpowiadających im statystyk w grupie **nbal**. Podobnie jest z parametrem AVA.post. Stąd niskie p-wartości testu Manna-Whitneya-Wilcozona dla hipotez alternatywnych o mniejszym parametrze położenia badanego parametru w grupie pacjentów **bal**.



Rysunek 3.2: Wykresy pudełkowe analizowanych parametrów serca: EF, AVA, Max, AVA.post i Max.post odpowiednio w porównaniu dla dwóch grup badanych pacjentów: **bal** oznaczonych TRUE i **nbal** oznaczonych FALSE.

3.4. Testowanie zbioru hipotez testem Manna-Whitneya-Wilcoxona

Dla wyliczonych w poprzednim rozdziale p-wartości poszczególnych pojedynczych testów można rozważać hipotezę główną:

H_0^{12345} : pacjenci z grup `bal` i `nbal` mają średnio **równe** parametry serca,

składającą się z omówionych już w poprzednim rozdziale hipotez H_0^i dla $i = 1, \dots, m$. Tak postawiona hipoteza, która będzie podlegała weryfikacji, powoduje rozważanie zagadnienia testowania zbioru hipotez, które omówiono w rozdziale 1.2. Do analizy testowania zbioru hipotez posłużą funkcje pakietu `multtest` opisane w rozdziale 2.2. Kod i wywołanie analizy rozważanego zagadnienia prezentuje się następująco:

Przypisanie p-wartości testów do jednego wektora o nazwie `p.wartosci`.

```
> p1<-wilcox.test(bal$AVA,nbal$AVA, alternative="less", paired=FALSE,
+ correct=FALSE, exact=TRUE)$p.value
> p2<-wilcox.test(bal$EF,nbal$EF, alternative="less", paired=FALSE,
+ correct=FALSE, exact=TRUE)$p.value
> p3<-wilcox.test(bal$Max,nbal$Max, alternative="less", paired=FALSE,
+ correct=FALSE, exact=TRUE)$p.value
> p4<-wilcox.test(bal$AVA.post,nbal$AVA.post, alternative="less", paired=FALSE,
+ correct=FALSE, exact=TRUE)$p.value
> p5<-wilcox.test(bal$Max.post,nbal$Max.post, alternative="less", paired=FALSE,
+ correct=FALSE, exact=TRUE)$p.value

> p.wartosci<-c(p1,p2,p3,p4,p5)
> p.wartosci
[1] 0.47839267 0.30855856 0.10002333 0.01929644 0.45118870
```

Użycie funkcji `mt.rawp2adjp` do wyliczenia dopasowanych p-wartości dla procedur Bonferroniego, Holma i Hochberga. Pakiet `R` zwraca, opisane w podrozdziale 2.2.1, wartości `adjp` i `index`. Wartość `adjp` to macierz, której kolumny odpowiadają dopasowanym p-wartościom dla zastosowanych procedur. Wartość `index` to wektor wskaźników oznaczający, którym z kolei początkowym p-wartościom odpowiadają wyliczone dopasowane p-wartości. Warto zwrócić uwagę, że dla procedur Bonferroniego i Holma, od trzeciej z kolei dopasowanej p-wartości, zwracane są wielkości bliskie 1, co oznacza nieodrzućenie odpowiadających im hipotez.

```
> mt.rawp2adjp(p.wartosci, proc=c("Bonferroni", "Holm", "Hochberg"))
$adjp
      rawp Bonferroni      Holm      Hochberg
[1,] 0.01929644 0.0964822 0.0964822 0.0964822
[2,] 0.10002333 0.5001167 0.4000933 0.4000933
[3,] 0.30855856 1.0000000 0.9256757 0.4783927
[4,] 0.45118870 1.0000000 0.9256757 0.4783927
[5,] 0.47839267 1.0000000 0.9256757 0.4783927

$index
[1] 4 3 2 5 1
```

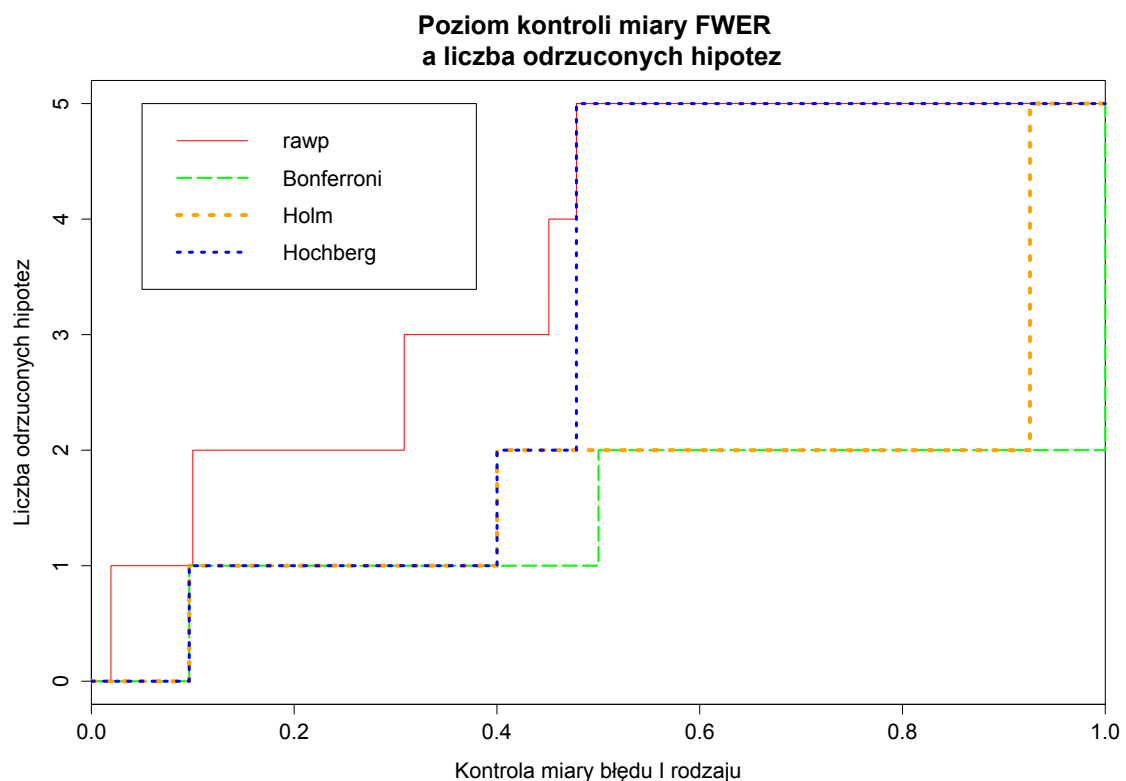
Wykorzystanie funkcji `mt.reject` do sprawdzenia, które hipotezy należy odrzucić w zagadnieniu testowania zbioru hipotez na poziomie kontroli miary błędu FWER $\alpha = 0.05$. Uzyskane wyniki przedstawione są jako wartości `r` i `which`, o których wspomniano w podrozdziale 2.2.2. W tym wypadku `r` to wektor, którego poszczególne wartości oznaczają ile hipotez zerowych zostało odrzuconych dla konkretnych procedur testowania. Widać, że w przypadku każdej z procedur nie została odrzucona ani jedna hipoteza zerowa pomimo, że w przypadku niezastosowania poprawki kontroli (wartość `rawp`) należałoby odrzucić jedną hipotezę zerową.

Hipoteza ta dotyczyła analizowanego parametru AVA.post. Wartość `which` to macierz wartości logicznych odzwierciedlająca, które hipotezy zerowe odrzucono. Kolumny odpowiadają procedurom, a wiersze kolejnym hipotezom uporządkowanym niemalejąco według odpowiadających im p-wartości.

```
> mt.reject(mt.rawp2adjp(p.wartosci, proc=c("Bonferroni", "Holm",
      "Hochberg"))$adjp, alpha=0.05)
$r
      rawp Bonferroni Holm Hochberg
0.05      1          0    0         0

$which
      rawp Bonferroni Holm Hochberg
[1,]  TRUE      FALSE FALSE  FALSE
[2,] FALSE      FALSE FALSE  FALSE
[3,] FALSE      FALSE FALSE  FALSE
[4,] FALSE      FALSE FALSE  FALSE
[5,] FALSE      FALSE FALSE  FALSE
```

Wywołanie funkcji `mt.plot2` do graficznego zaprezentowania zależności między liczbą odrzuconych hipotez zerowych, dla poszczególnych procedur, a poziomem kontroli błędu FWER. Kod ujawniono w Dodatku A. Efekt został ukazany na Rysunku 3.3.



Rysunek 3.3: Poziom kontroli miary FWER a liczba odrzuconych hipotez dla procedur Bonferroni (na zielono), Holma (na żółto) i Hochberga (na niebiesko). Czerwonym kolorem ukazano odrzucenia dla niezastosowanej poprawki poziomów istotności pojedynczych testów w przypadku zagadnienia testowania zbioru hipotez.

3.5. Wyniki i wnioski

Przeprowadzona analiza wykazuje, że nie ma podstaw do odrzucenia ani jednej z hipotez w rozważanym zagadnieniu testowania zbioru hipotez dla żadnej, nawet najmniej restrykcyjnej, rozważanej procedury kontroli FWER, co oznacza, że nie wykazano, aby którakolwiek z badanych metod przeprowadzania zabiegu w przypadku zagrożenia życia pacjenta, jakim jest zawał, była skuteczniejsza. Obie metody przeprowadzano na grupie pacjentów o równie zaburzonych parametrach serca przed zabiegiem (parametry EF,AVA i Max) i obie metody przywróciły pacjentów do stanu równie stabilnego po zabiegu (parametry AVA.post i Max.post). Nie ma podstaw do faworyzowania żadnej z metod. Jedyną cechą przeważającą za metodą balonikowania jest silna inwazyjność metody bypassów, która jest metodą przeprowadzania zabiegu na otwartym sercu.

Podsumowanie

W pracy przedstawiono teorię weryfikacji hipotez statystycznych i teorię testowania zbioru hipotez. Ukazano główne problemy tego zagadnienia i związane z nimi rozwiązania w postaci klasycznych procedur testowania zbioru hipotez. Wybrano, opisano i porównano 3 procedury testowania kontrolujące prawdopodobieństwo popełnienia co najmniej jednego błędu w procesie testowania hipotez, polegającego na błędnym odrzuceniu prawdziwej hipotezy zerowej. Omówione procedury Bonferroniego, Holma i Hochberga zastosowano do danych medycznych w celu porównania, która z dwóch metod przeprowadzania zabiegu jest skuteczniejsza, co było celem całej pracy.

Dane rozważane w pracy podzielono na dwa podzbiory pacjentów, w zależności od przeprowadzonej metody zabiegu. Porównano parametry pracy serca takie jak frakcja wyrzutowa serca, powierzchnia przekroju aorty czy maksymalny gradient ciśnień w dwóch grupach pacjentów przed i po zabiegu. Zrezygnowano z poddawania analizie parametru określającego średni gradient ciśnień ze względu na jego silną zależność od maksymalnego gradientu ciśnień, zarówno przed jak i po zabiegu.

Przedstawiono opis i zastosowania funkcji dostępnych w pakietach `stats` i `multtest` w celu wykorzystania ich do analizy. Użyto również graficznego pakietu `ggplot2` do wytworzenia grafik i wykresów przedstawionych w pracy.

Szczegółowo opisano i przeprowadzono testy Manna-Whitneya-Wilcoxona, dla których następnie wygenerowano dopasowane p-wartości przy pomocy omówionych w pracy funkcji pakietu statystycznego `R`. Powołując się na teorię testowania zbioru hipotez, przeprowadzono analizę dwóch metod zabiegów, uwzględniając poprawkę poziomowi istotności dla każdej z rozważanych hipotez zerowych dotyczących wielkości parametrów położenia parametrów serca w dwóch badanych grupach pacjentów.

Dzięki analizie nie stwierdzono dominacji żadnej z metod nad drugą, jeżeli chodzi o porównanie parametrów pracy serca przed zabiegiem i po nim. Wykorzystano 3 procedury testowania z racji ich różnych poziomów mocy, czyli zdolności do odrzucania fałszywych hipotez. Jednak, na badanym poziomie miary błędu $\alpha = 0.05$, nie stwierdzono różnic w wynikach dla żadnej z nich.

Dodatek A

Kody pakietu \mathcal{R} użyte w pracy

W tym rozdziale zamieszczone są kody pakietu \mathcal{R} , których użyto do analizy danych w pracy oraz do wygenerowania wykresów zamieszczonych w pracy.

Obróbka danych

Wczytanie danych i mała korekta. Usunięto obserwacje dotyczące jednego z pacjentów (nr 26), gdyż nie zawierały żadnych danych po zabiegu. Najprawdopodobniej pacjent zmarł. Argument `dec=","` zaznacza, że w danych pierwotnych kropka dziesiętna była oznaczana znakiem `","` i zostanie zastąpiona znakiem `."`, który jest rozpoznawany przez \mathcal{R} .

```
> dane = read.csv("C:/sciezka_do_pliku_z_danymi/zawaly.csv", sep=";", dec=",")
> dane = dane[-26,]
```

Podział całego zbioru danych na dwa podzbiory: pacjentów potraktowanych balonikowaniem i potraktowanych metodą bypassów.

```
indeksy = dane[,1] == 1
bal = dane[indeksy,]
nbal = dane[!indeksy,]
```

Analiza danych

Wyliczenie statystyk agregujących ukazanych w Tabeli 3.2.

```
summary(bal[,c(-1,-5,-8-,9)])
summary(nbal[,c(-1,-5,-8-,9)])
```

Zastosowanie testu kombinatorycznego Manna-Whitneya-Wilcozona dla dwóch próbek dla zmiennych EF, AVA, Max, AVA.post i Max.post. Próbkę brane są z podzbiorów: bal i nbal. Tabela 3.1.

```
wilcox.test(bal$AVA,nbal$AVA, alternative="less", paired=FALSE,
            correct=FALSE, exact=TRUE)
wilcox.test(bal$EF,nbal$EF, alternative="less", paired=FALSE,
            correct=FALSE, exact=TRUE)
wilcox.test(bal$Max,nbal$Max, alternative="less", paired=FALSE,
            correct=FALSE, exact=TRUE)
wilcox.test(bal$AVA.post,nbal$AVA.post, alternative="less", paired=FALSE,
            correct=FALSE, exact=TRUE)
wilcox.test(bal$Max.post,nbal$Max.post, alternative="less", paired=FALSE,
            correct=FALSE, exact=TRUE)
```

Przykład z podrozdziału 2.2.4 wraz z Rysunkiem 2.1.

```
przykladP <- c(0.018,0.010,0.023,0.11)
mt.rawp2adjp(przykladP, proc=c("Holm", "Hochberg"))
mt.reject(mt.rawp2adjp(przykladP, proc=c("Holm", "Hochberg"))$adjp, alpha=0.05)

mt.plot2(mt.rawp2adjp(przykladP, proc=c("Holm", "Hochberg"))$adjp,
  plottype="rvsa", logscale=FALSE,
  alpha=seq(0, 0.15, length = 10000),
  proc=c("rawp","Holm", "Hochberg"), leg=c(0.1,1.5),
  col=c("red","orange","blue"), main="Poziom_kontroli_miary_FWER
  \n_a_liczba_odrzuconych_hipotez", lty=c(1:3), lwd=c(1,2,3),
  font.main=2, xlab = "Kontrola_miary_bledu_I_rodzaju",
  ylab = "Liczba_odrzuconych_hipotez",)
```

Testowanie zbioru hipotez testem Manna-Whitneya-Wilcoxon. Rozdział 3.4.

```
> p1<-wilcox.test(bal$AVA,nbal$AVA, alternative="less", paired=FALSE,
  correct=FALSE, exact=TRUE)$p.value
> p2<-wilcox.test(bal$EF,nbal$EF, alternative="less", paired=FALSE,
  correct=FALSE, exact=TRUE)$p.value
> p3<-wilcox.test(bal$Max,nbal$Max, alternative="less", paired=FALSE,
  correct=FALSE, exact=TRUE)$p.value
> p4<-wilcox.test(bal$AVA.post,nbal$AVA.post, alternative="less", paired=FALSE,
  correct=FALSE, exact=TRUE)$p.value
> p5<-wilcox.test(bal$Max.post,nbal$Max.post, alternative="less", paired=FALSE,
  correct=FALSE, exact=TRUE)$p.value

> p.wartosci<-c(p1,p2,p3,p4,p5)
> p.wartosci

> mt.rawp2adjp(p.wartosci, proc=c("Bonferroni", "Holm", "Hochberg"))
> mt.reject(mt.rawp2adjp(p.wartosci, proc=c("Bonferroni", "Holm",
  "Hochberg"))$adjp, alpha=0.05)
```

Wywołanie funkcji mt.plot2. Rysunek 3.3.

```
mt.plot2(mt.rawp2adjp(prown, proc=c("Holm", "Hochberg"))$adjp,
  plottype="rvsa", logscale=FALSE, alpha=seq(0, 1, length = 1000),
  proc=c("rawp","Holm", "Hochberg"), leg=c(0.05,3),
  col=c("red","orange","blue"),
  main="Poziom_kontroli_miary_FWER\n_a_liczba_odrzuconych_hipotez",
  lty=c(1:3), lwd=c(1,2,3), font.main=2,
  xlab = "Kontrola_miary_bledu_I_rodzaju",
  ylab = "Liczba_odrzuconych_hipotez",)
```

Przypisanie funkcji mt.plot2 i wymuszone przypisanie funkcji legenda w celu wygenerowania dowolnych podpisów pod osiami na Rysunku 3.3.

```
mt.plot2<- mt.plot
legenda <- multtest::mt.legend
```

Wprowadzone poprawki do funkcji mt.plot2 w porównaniu do funkcji mt.plot. Z wymienionych poniżej linii kodu funkcji mt.plot2, aby usunąć z góry zadawane podpisy pod osiami, zmieniono pierwotne linie na

```
fix(mt.plot2)
[10] matplot(alpha, r,
[11]           type = "l", ...)

[16]matplot(1:m, spval,
[17]           type = "l", ...)
```

```
[22]matplot(teststat, adjp,
[23]          type = "p", ...)

[25]matplot(teststat, -log(adjp, 10),
[26]          type = "p",

[32]matplot(1:m, adjp,

[35]matplot(1:m, -log(adjp, 10),
```

oraz, aby móc korzystać z funkcji `mt.legend` niedostępnej dla użytkowników zamieniono w kodzie funkcję `mt.legend` na jej dokładną kopię, funkcję `legenda`. Linie w kodzie funkcji `mt.plot2` zamieniono na:

```
[12]legenda(leg[1], leg[2], proc, ...)
[18]legenda(leg[1], leg[2], proc, ...)
[28]legenda(leg[1], leg[2], proc, ...)
[37]legenda(leg[1], leg[2], proc, ...)
```

Funkcja `panel.pearson`.

```
library(graphics)

panel.pearson <- function(x, y, ...) {
  horizontal <- (par("usr")[1] + par("usr")[2]) / 2;
  vertical <- (par("usr")[3] + par("usr")[4]) / 2;
  text(horizontal, vertical,
        format(abs(cor(x,y, use = "pairwise.complete.obs")), digits=2), cex=4) }
```

Wygenerowanie macierzy korelacji i wykresów rozrzutu z Rysunku 3.1.

```
dane.cor = dane[,c(-1,-9)]
pairs(~AVA+EF+Max+Mean+Max.post+Mean.post+AVA.post,
      data=dane.cor, upper.panel=panel.pearson, pch=21, cex=0.8,
      font.labels=2, cex.labels=1.9,
      main="Macierz_korelacji", cex.main=2)
```

Wykresy

Wykres przedstawiający prawdopodobieństwo popełnienia co najmniej jednego błędu I rodzaju w zależności od liczby rozważanych hipotez. Rysunek 1.1.

```
x=c(1:200)
names(x)="Liczba_rozważanych_hipotez"
y1=1-(0.95)^x
y2=1-(0.98)^x
y3=1-(0.99)^x
y4=1-(0.9)^x
Z=data.frame(x,y1,y2,y3,y4)
names(Z)=c("Liczba_rozważanych_hipotez",
           "Prawdopodobieństwo_popełnienia_błędu_I_rodzaju")

ggplot()+
  layer(data=Z, mapping=aes(Liczba_rozważanych_hipotez,
                           Prawdopodobieństwo_popełnienia_błędu_I_rodzaju), geom="line",
        col="blue", lwd=1)+
  layer(data=Z, mapping=aes(x,y2), geom="line", col="orange", lwd=1)+
  layer(data=Z, mapping=aes(x,y3), geom="line", col="red", lwd=1)+
  layer(data=Z, mapping=aes(x,y4), geom="line", col="green", lwd=1)+
  annotate("text", label = "0.01", x = 150, y = 0.2, size = 5,
         colour = "red")+
  annotate("text", label = "0.10", x = 150, y = 0.5, size = 5,
```

```

colour = "green")+
annotate("text", label = "0.05", x = 150, y = 0.4, size = 5,
colour = "blue")+
annotate("text", label = "0.02", x = 150, y = 0.3, size = 5,
colour = "orange")+
annotate("text", label = "Poziomy istotności", x = 150, y = 0.6, size = 5,
colour = "black")

```

Wykresy pudełkowe uzyskane przy użyciu pakietu ggplot2. Rysunek 3.2.

```

dane123 = dane
dane123[,1] = as.logical(dane123[,1])
library(ggplot2)

wykres1 <- qplot(xlab="Typ zabiegu", TypZabiegu, EF, data=dane123,
                ylab="EF", main="Wykresy pudełkowe dla parametru EF")+
  geom_boxplot(width=0.3, alpha=0.99)+
  geom_jitter(size=0.01)
wykres1

wykres2 <- qplot(xlab="Typ zabiegu", TypZabiegu, AVA, data=dane123,
                ylab="AVA", main="Wykresy pudełkowe dla parametru AVA")+
  geom_boxplot(width=0.3, alpha=0.99)+
  geom_jitter(size=0.01, colour="blue")
wykres2

wykres3 <- qplot(xlab="Typ zabiegu", TypZabiegu, Max, data=dane123,
                ylab="Max", main="Wykresy pudełkowe dla parametru Max")+
  geom_boxplot(width=0.3, alpha=0.99)+
  geom_jitter(size=0.01, colour="blue")
wykres3

wykres5 <- qplot(xlab="Typ zabiegu", TypZabiegu, AVA.post, data=dane123,
                ylab="AVA.post",
                main="Wykresy pudełkowe dla parametru AVA.post")+
  geom_boxplot(width=0.3, alpha=0.99)+
  geom_jitter(size=0.01, colour="blue")
wykres5

wykres6 <- qplot(xlab="Typ zabiegu", TypZabiegu, Max.post, data=dane123,
                ylab="Max.post",
                main="Wykresy pudełkowe dla parametru Max.post")+
  geom_boxplot(width=0.3, alpha=0.99)+
  geom_jitter(size=0.01, colour="blue")
wykres6

```


Bibliografia

- [1] Benjamini Y., Hochberg Y., *Controlling the false discovery rate: a practical and powerful approach to multiple testing*, J. R. Statist. Soc. B, 1995.
- [2] Biecek P., *Testowanie zbioru hipotez z zadaną relacją hierarchii wraz z przykładami zastosowań w genetyce*, Rozprawa doktorska, Wydawnictwo Politechniki Wrocławskiej, Wrocław 2007.
- [3] Bonferroni C., *Teoria statistica delle classi e calcolo delle probabilità*, Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze, 1936.
- [4] Bretz F., Hothorn T., Westfall P., *Multiple Comparisons Using R*, Chapman & Hall, 2010.
- [5] Dudoit S., Shaffer J., Boldrick J., *Multiple hypothesis testing in microarray experiments*, Statistical Science 18 (1): 71-103, 2003.
- [6] Dudoit S., Ge Y., *Bioconductor's multtest package*, Opis pakietu multtest, 2005, <http://www.bioconductor.org/packages/release/bioc/vignettes/multtest/inst/doc/multtest.pdf>.
- [7] Dudziński M., Furmańczyk K., *Procedury jednoczesnego testowania wielu hipotez i ich zastosowania w analizie mikromacierzy DNA*, Matematyka Stosowana 8, 2007.
- [8] Dudziński M., Furmańczyk K., *Wykorzystanie pakietu R w zagadnieniach dotyczących procedur jednoczesnego testowania wielu hipotez*, http://iks_pn.sggw.pl/z35/art1.pdf.
- [9] Hochberg Y., Tamhane A., *Multiple Comparison Procedures*, New York, 1987.
- [10] Holm S., *A simple sequentially rejective multiple test procedure*, Scandinavian Journal of Statistics, 1979.
- [11] Jakubowski J., Sztencel R., *Wstęp do teorii prawdopodobieństwa*, SCRIPT, Wydanie IV, 2010.
- [12] Lehmann E., Romano J., *Generalizations of the familywise error rate*, Ann. Statist. 33, 2005.
- [13] Niederlaender E., *Causes of death in the EU - Statistics in focus*, 10/2006.
- [14] Niemirow W., Skrypt do przedmiotu *Statystyka*, 2011, <http://www-users.mat.umk.pl/~wniem/Statystyka/Statystyka.pdf>.

- [15] Pollard K., Gilbert H., Ge Y., Taylor S., Dudoit S., *multtest: Resampling-based multiple hypothesis testing*. R package version 2.16.0.
- [16] R Core Team, *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Wiedeń 2013, ISBN 3-900051-07-0, <http://www.R-project.org/>.
- [17] Szczeklik A. (red.), *Choroby wewnętrzne. Przyczyny, rozpoznanie i leczenie. Tom 1*, Wydawnictwo: Medycyna Praktyczna, 2005.
- [18] Zieliński R., *Siedem wykładów wprowadzających do statystyki matematycznej*, PWN, Warszawa 1990.