

**Uniwersytet Warszawski**  
Wydział Matematyki, Informatyki i Mechaniki

**Anna Kościńska, Karolina Orczykowska**

Nr albumu: 311725, 312073

**Testy diagnostyczne dla trzech  
grup. Zastosowanie pakietu  
DiagTest3Grp w analizie danych  
medycznych**

**Praca licencjacka  
na kierunku MATEMATYKA  
w zakresie JEDNOCZESNYCH STUDIÓW  
EKONOMICZNO-MATEMATYCZNYCH**

Praca wykonana pod kierunkiem  
**dra inż. Przemysława Biecka**  
Instytut Matematyki Stosowanej i Mechaniki UW

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**

W niniejszej pracy podjęta została próba wskazania klasyfikatora o najlepszej zdolności różnicującej pacjentów, których uprzednio podzielono na trzy grupy pod względem zaawansowania choroby. W tym celu zastosowano pakiet `DiagTest3Grp` w programie `R` na przykładzie danych dotyczących zaburzeń psychicznych.

## **Słowa kluczowe**

czułość, swoistość, test diagnostyczny, ROC, VUS, indeks Youdena, test Shapiro–Wilka, `DiagTest3Grp`

## **Dziedzina pracy (kody wg programu Socrates-Erasmus)**

11.2 Statystyka

## **Klasyfikacja tematyczna**

92C50–Biology and other natural sciences–Physiological, cellular and medical topics–Medical applications (general)

62P10–Statistics–Applications–Applications to biology and medical sciences

## **Tytuł pracy w języku angielskim**

Diagnostic tests for the three groups. Application of the package `DiagTest3Grp` in the analysis of medical data.



# Spis treści

<b>Wprowadzenie</b> . . . . .	11
<b>1. Teoria</b> . . . . .	13
1.1. Krzywa ROC jako ocena jakości klasyfikatora — przypadek binarny . . . . .	13
1.2. Rozszerzenie krzywej ROC na przypadek testów dla trzech grup — VUS . . . . .	15
1.2.1. Estymacja VUS . . . . .	17
1.3. Indeks Youdena . . . . .	18
1.3.1. Estymacja parametryczna . . . . .	19
1.3.2. Estymacja nieparametryczna . . . . .	20
1.3.3. Parametryczne przedziały ufności . . . . .	21
1.3.4. Nieparametryczne przedziały ufności . . . . .	22
<b>2. Funkcje pakietu R do analizy klasyfikatorów: biblioteka DiagTest3Grp</b> . . . . .	25
2.1. Funkcje Youden3Grp i VUS . . . . .	25
2.2. Inne funkcje pakietu DiagTest3Grp . . . . .	28
<b>3. Analiza danych medycznych</b> . . . . .	31
3.1. Opis zbioru danych . . . . .	31
3.2. Badanie rozkładu . . . . .	33
3.2.1. Test Shapiro–Wilka . . . . .	33
3.2.2. Transformacja Boxa-Coxa . . . . .	35
3.3. Porównanie własności pojedynczych markerów . . . . .	41
3.4. Analiza wszystkich markerów — porównanie . . . . .	49
<b>4. Podsumowanie</b> . . . . .	53
<b>A. Dowód wieloczynnikowej metody delta</b> . . . . .	55
<b>B. Częściowe pochodne <math>t_-^*</math>, <math>t_+^*</math> i <math>J</math></b> . . . . .	57
<b>C. Wykresy pudełkowe i rozproszenia VUS</b> . . . . .	59
<b>Bibliografia</b> . . . . .	63



# Spis rysunków

1.1. Przykład krzywej ROC . . . . .	15
1.2. Przykład wykresu powierzchni ROC. . . . .	16
3.1. Zależność zmiennych <i>MANSA</i> i <i>BPRS.Average</i> od miasta badania . . . . .	32
3.2. Histogramy badanych markerów. . . . .	34
3.3. Wykres Q-Q dla kolejnych zmiennych . . . . .	36
3.4. Wykres zależności wartości log-prawdopodobieństwa od wartości parametru $\lambda$ dla zmiennej <i>BPRS.Negative</i> . . . . .	38
3.5. Wykresy Q-Q dla trzech grup zmiennej <i>BPRS.Negative</i> (od lewej: $D^-$ , $D^0$ i $D^+$ ) . . . . .	39
3.6. Histogramy zmiennej <i>BPRS.Negative</i> od lewej: przekształconej transforacją Boxa-Coxa, pierwotnej . . . . .	40
3.7. Wykres rozproszenia oraz wykres pudełkowy <i>BPRS.Depression</i> . . . . .	43
3.8. Wartość wyestymowanej miary VUS przy pomocy metody parametrycznej i nieparametrycznej dla wszystkich markerów. . . . .	44
3.9. Analiza wszystkich markerów w kontekście rozszerzonego indeksu Youdena . . . . .	48
3.10. Mapa ciepła <i>p-value</i> dopasowanego dla sparowanych klasyfikatorów: VUS. . . . .	52
3.11. Mapa ciepła <i>p-value</i> dopasowanego dla sparowanych klasyfikatorów: indeks Youdena. . . . .	52
C.1. Wykres rozproszenia oraz wykres pudełkowy <i>BPRS.Negative</i> . . . . .	59
C.2. Wykres rozproszenia oraz wykres pudełkowy <i>BPRS.Average</i> . . . . .	60
C.3. Wykres rozproszenia oraz wykres pudełkowy <i>BPRS.Maniac</i> . . . . .	60
C.4. Wykres rozproszenia oraz wykres pudełkowy <i>BPRS.Positive</i> . . . . .	61





# Spis tabel

2.1. Opis argumentów funkcji VUS . . . . .	26
2.2. Opis argumentów funkcji Youden3Grp . . . . .	27
2.3. Opis argumentów funkcji DiagTest3Grp.Test . . . . .	29
3.1. Zmienne zbioru danych EDEN . . . . .	31
3.2. Średnie w każdej z grup dla kolejnych zmiennych . . . . .	33
3.3. Wartości parametru $\lambda$ transformacji Boxa–Coxa dla kolejnych zmiennych oraz wyniki testu Shapiro — Wilka dla trzech grup. . . . .	40
3.4. Wyniki estymacji miary VUS . . . . .	42
3.5. Szczegółowa analiza poszczególnych zmiennych: rozszerzony indeks Youdena . . . . .	47
3.6. Statystyka testowa $Z$ oraz $p$ -value dla porównywania parami objętości pod powierzchnią ROC . . . . .	50
3.7. Statystyka testowa $Z$ oraz $p$ -value dla porównywania parami rozszerzonego indeksu Youdena . . . . .	51



# Udział w przygotowaniu pracy

Niniejsza praca została napisana przez dwie osoby: Karolinę Orczykowską oraz Annę Kościńską. Praca była tworzona wspólnie, ale można wskazać osoby, które przyczyniły się w większym stopniu do napisania poszczególnych rozdziałów. Dla Karoliny Orczykowskiej są to podrozdziały: 1.1, 1.2, 1.4, rozdział 2, podrozdział 3.4, Dodatek C; dla Anny Kościńskiej: 1.3, 3.1, 3.2., Podsumowanie, Dodatek A i Dodatek B. Wprowadzenie i podrozdział 3.3 były pisane wspólnie.



# Wprowadzenie

Tradycyjnie, testy diagnostyczne są podstawowym sposobem na oszacowanie szansy, że dany badany jest chory. Na ich podstawie można wysnuć przypuszczenie o możliwej diagnozie, jednakże ze względu na to, iż nie dają stuprocentowej pewności, ich wyniki należy potwierdzić innymi, bardziej specjalistycznymi badaniami. Testy te są często proste do przeprowadzenia oraz łatwo dostępne, dlatego stanowią pierwszy krok ku wykryciu potencjalnej choroby. Pomimo że, przyczyniają się do postawienia hipotezy o stanie zdrowia pacjenta, nie stanowią ostatecznej diagnozy — są jedynie jej przesłankami. W przypadku wielu chorób bardzo duży wpływ na całkowite wyleczenie ma wczesne wykrycie ich symptomów. Możliwe jest wtedy zastosowanie skuteczniejszej, mniej inwazyjnej terapii. Dzięki powszechności testów diagnostycznych, możliwe jest stosunkowo szybkie ich przeprowadzenie oraz w razie konieczności, podjęcie dalszych kroków.

Wśród testów diagnostycznych wyróżnia się m.in. testy binarne, zwracające wynik pozytywny lub negatywny, oraz testy dla trzech grup. Podstawową różnicą pomiędzy nimi jest to, że testy dla trzech grup umożliwiają otrzymanie wyniku „pośredniego”, który odnosi się do wczesnego stadium choroby. Zda się więc, że w pewnych przypadkach, takich jak np. choroba Alzheimera czy przy wykrywaniu nowotworów, są one bardzo cennym narzędziem. Ze względu na ich znacznie mniejszą popularność oraz niewątpliwą użyteczność, warto jest je lepiej poznać i rozpowszechniać. W pracy tej zajęto się nie tylko wyjaśnieniem teoretycznych aspektów rozszerzenia testów binarnych o trzecią grupę, ale również pokazaniem na konkretnym przykładzie metodologii sprawdzania, wyniki którego z markerów mogłyby posłużyć jako test diagnostyczny przy danej chorobie.

W pracy tej jako test przyjmuje się wartości poszczególnego markera dla danej osoby. Następnie, na podstawie wcześniej zdefiniowanych trzech grup — zdrowych, chorych i tych w początkowym stadium choroby, sprawdza się, czy dany klasyfikator podobnie rozróżnia badanych — jego wartości można podzielić na trzy przedziały takie, że każdy z nich odpowiada jednej z powyższych grup. Istotne jest odpowiednie wybranie punktów rozdzielających te przedziały.

Często zdarza się, że wartości pierwotnej zmiennej rozdzielającej pacjentów do trzech grup są teoretycznie określone, ale ich praktyczne odnalezienie (czyli przebadanie chorego) jest trudne lub obciążające zdrowie pacjenta. Dlatego też ważnym jest wyselekcjonowanie takich parametrów, których określenie wartości nie sprawia problemów, a które są czułe na chorobę — tzn ich wartości zmieniają się wraz z rozwojem choroby w podobny sposób u każdej osoby.

Każdy test charakteryzuje się innym poziomem poprawności diagnostycznej, czyli zdolnością do właściwego rozróżnienia badanych do każdej z grup. W diagnostyce medycznej poziom ten bazuje na dwóch miarach — czułości i swoistości. Pary punktów tych wartości wyznaczają w przypadku binarnym krzywą ROC. Pole pod jej wykresem (ozn. AUC) charakteryzuje prawdopodobieństwo poprawnego przypisania pacjenta do odpowiedniej grupy. Analogicznie, dla trzech grup krzywa ROC rozszerza się do powierzchni, a AUC do objętości

pod tą powierzchnią — miary VUS. Innym wskaźnikiem poprawności testu diagnostycznego jest indeks Youdena. Jest on kombinacją czułości i swoistości. Można go również rozszerzyć na przypadek trzech grup.

Rozpatrywanie testu dla trzech grup warto rozpocząć od zbadania powyższych miar. Mogą one wskazać nie tylko z jakim prawdopodobieństwem wyniki testu są prawidłowe, ale też optymalne punkty rozdzielające wartości rozpatrywanego markera dla pacjentów chorych, zdrowych i we wczesnej fazie choroby.

Celem tej pracy jest próba wyselekcjonowania takich markerów spośród dostępnych, które charakteryzują się najlepszymi własnościami diagnostycznymi. Zbiorem danych, na których przeprowadzana jest analiza, są dane rzeczywiste dotyczące chorób psychicznych. Pacjenci dzieleni są na podstawie wyniku testu określającego ich jakość życia. Ponieważ jest to zmienna trudna do obiektywnego zmierzenia, poszukiwane są markery, które umożliwiają w podobny sposób grupowanie badanych. Dostępnymi klasyfikatorami są oceny objawów psychicznych takich jak lęk, depresja, omamy oraz zachowania nietypowe, według powszechnie stosowanej skali BPRS.

Wszystkie obliczenia wykonane zostały na podstawie stworzonego w tym celu pakietu `DiagTest3Grp` w programie R.

Struktura pracy jest następująca. W Rozdziale 1 znajdują się teoretyczne podstawy analizy poprawności testów diagnostycznych. W Rozdziale 2 opisane zostały użyte funkcje z pakietu `DiagTest3Grp`. Rozdział 3 to wyniki przeprowadzonych analiz na danych rzeczywistych. Podsumowanie znajduje się w Rozdziale 4.

# Rozdział 1

## Teoria

W rozdziale tym przedstawione zostaną teoretyczne zagadnienia związane z analizą klasyfikatorów wykorzystywanych w testach diagnostycznych. Znajdą się w nim szczegółowe opisy takich pojęć, jak krzywa ROC (ang. *Receiver Operating Characteristic*) — podrozdział 1.1, VUS (ang. *Volume Under Surface*) — podrozdział 1.2, rozszerzony indeks Youdena — podrozdział 1.3.

### 1.1. Krzywa ROC jako ocena jakości klasyfikatora — przypadek binarny

Jednym z narzędzi służących do oceny poprawności klasyfikatora w diagnostyce medycznej jest krzywa ROC (ang. *Receiver Operating Characteristic*). Krzywa ROC stosowana jest jedynie dla testów biomedycznych typu binarnego — tzn. zwracających jedynie dwa wyniki: pozytywny (stan chorobowy) oraz negatywny (brak symptomów choroby).

Krzywa ta wyznaczana jest przez dwa współczynniki: *czułość* (ang. *sensitivity*) oraz *swobodność* (ang. *specificity*).

Do zdefiniowania tych terminów pomocne jest określenie stanu klasyfikatora na podstawie stanu faktycznego i wskazań testu (z [8]):

TP (ang. *true positive*) — liczba poprawnie sklasyfikowanych pacjentów chorych,

FN (ang. *false negative*) — liczba błędnie sklasyfikowanych pacjentów chorych,

TN (ang. *true negative*) — liczba poprawnie sklasyfikowanych pacjentów zdrowych,

FP (ang. *false positive*) — liczba błędnie sklasyfikowanych pacjentów zdrowych.

**Definicja 1.1.1** *Czułością danego testu określa się prawdopodobieństwo, że osoba chora zostanie zakwalifikowana do grupy chorych. Równoważnie, jest to miara określająca, jaki w przybliżeniu procent osób chorych zostanie poprawnie rozpoznany.*

W literaturze [8] oznacza się ją jako TPR (ang. *True Positive Rate*). W dalszej części pracy oznaczana jest jako  $Se$ .

Dla danego testu i znanych TP, FN, TN, FP miarę  $Se$  można przybliżać za pomocą wzoru:

$$Se \approx \frac{TP}{TP + FN}. \quad (1.1)$$

Wrażliwość bliska 1 sygnalizuje, że klasyfikator poprawnie rozpoznaje pacjentów chorych — małe prawdopodobieństwo popełnienia błędu pierwszego rodzaju. W takim przypadku mówi się, że test jest wrażliwy na wykrycie choroby.

**Definicja 1.1.2** *Swoistość określa prawdopodobieństwo, że osobę zdrową test zakwalifikuje do grupy zdrowej. Równoważnie, jest to miara określająca, jaki w przybliżeniu procent osób zdrowych zostanie poprawnie rozpoznany.*

W literaturze [8] wprowadza się wielkość FPR (ang. *False Positive Rate*) czyli prawdopodobieństwo uzyskania fałszywych przewidywanych sukcesów. FPR i swoistość łączy prosta zależność: *swoistość* = 1 - FPR. W pracy tej swoistość oznaczana jest jako  $Sp$ .

Podobnie jak w przypadku  $Se$ , miarę  $Sp$  można aproksymować jako:

$$Sp \approx \frac{TN}{FP + TN} = 1 - FPR = 1 - \frac{FP}{FP + TN}. \quad (1.2)$$

Do wyznaczenia ROC stosuje się wartość *1-swoistość*. Dla testów diagnostycznych pożądana jest niska jej wartość (tzn. bliska 0). Daje to pewność, że relatywnie niewielu zdrowych badanych zostaje zidentyfikowanych jako chorzy — małe prawdopodobieństwo popełnienia błędu drugiego rodzaju.

Zatem pożądanym testem, czy też niemalże idealnie poprawnie działającym klasyfikatorem (biomarker), charakteryzuje się wysokimi wartościami *czułości* i *swoistości* (odpowiednio niską wartością *1-swoistość*). Dobry test to taki, który minimalizuje liczbę błędów — czyli liczby FN i FP dla dowolnej próby. Dla różnych testów jednak jest tak, że popełnianie któregoś z tych błędów ma gorsze skutki — np. podczas klasyfikowania pacjentów do grup ryzyka większym błędem będzie zakwalifikowanie chorej osoby do grupy zdrowych niż odwrotnie.

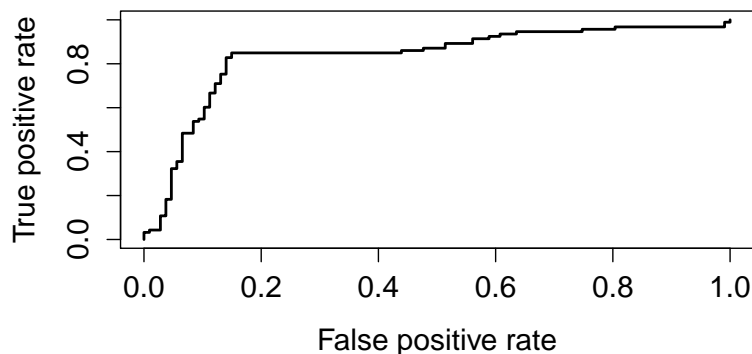
Krzywa ROC jest dwuwymiarowym wykresem z *czułością* na osi Y i wartością *1-swoistość* na osi X wyznaczającym wszystkie możliwe kombinacje tych zmiennych dla testu. Krzywa ta istnieje, ponieważ możliwym jest uzyskanie nieskończonej liczby wydzielonych klasyfikatorów, dla których obie powyższe cechy brane są z różnymi wagami. Obraz tej krzywej mieści się w kwadracie jednostkowym pierwszej ćwiartki układu współrzędnych. Charakterystycznymi punktami są: (0,0) — test daje zawsze wynik negatywny; (1,1) — test zawsze daje wynik pozytywny; (0,1) — idealna klasyfikacja; punkty typu (x,x) dla  $x \in (0,1)$  oznaczają pełną losowość wyniku. Przykładowa krzywa ROC dla testu diagnostycznego na wirus HIV znajduje się na rysunku 1.1.

Z testami zwracającymi wynik w postaci binarnej wiąże się pojęcie powierzchni pod krzywą ROC (ang. *Area under ROC curve*, AUC). Jest to dość popularne narzędzie do oceny i porównywania między sobą modeli klasyfikacyjnych. AUC można interpretować jako prawdopodobieństwo, że dany klasyfikator nada wyższą rangę losowo wybranemu przypadkowi pozytywnemu aniżeli losowo wybranemu przypadkowi z grupy, o której wiadomo, że nie występuje jej dany symptom chorobowy. Im większą wartość ma AUC, tym dany test uznaje się za lepszy. Perfekcyjny test charakteryzuje się AUC równym 1, podczas gdy kompletnie bezużyteczny — całkowicie losowy, ma AUC równe  $\frac{1}{2}$ . AUC wykorzystywanych w diagnostyce medycznej testów mieści się w przedziale od  $\frac{1}{2}$  do 1. W literaturze [2] spotyka się, że testy dzieli się na testy:

- wysokiej dokładności  $AUC \geq 0.9$ ,
- umiarkowanej dokładności  $0.7 \leq AUC < 0.9$ ,



**Rysunek 1.1:** Przykładowy wykres przedstawiający krzywą ROC dla jednego z klasyfikatorów używanych w testach wykrywających wirus HIV. Na osi poziomej znajdują się wartości  $1 - Sp$ , a na pionowej odpowiadające im wartości  $Se$ . Krzywa ta wyznacza wszystkie możliwe kombinacje tych punktów dla danego testu. Krzywa została wykonana przy pomocy pakietu [20].



- niskiej dokładności  $0.5 \leq AUC < 0.7$ ,
- o losowym wyniku  $AUC < 0.5$ .

Obie powyższe miary okazują się bardzo pomocne do porównywania diagnostycznej dokładności dwóch lub więcej testów, czyli w podejmowaniu decyzji o tym, który z testów okazuje się poprawniejszy do konkretnej sytuacji.

Krzywa ROC oraz AUC stosowane są jedynie w przypadkach wykorzystania dwóch klas, jakkolwiek analiza ROC może być rozszerzona na teorię wielowymiarowych problemów.

## 1.2. Rozszerzenie krzywej ROC na przypadek testów dla trzech grup — VUS

Praca ta rozważa jednak przypadek testów, dla których otrzymywana jest klasyfikacja nie na dwie, ale na trzy grupy. Rozróżnia się więc: wczesne stadium choroby, zaawansowane stadium choroby oraz brak symptomów. Umieszczenie dodatkowego stanu do sklasyfikowania przez dany test nieco komplikuje metody, które wykorzystuje się do tworzenia najlepiej dyskryminującego testu. Komplikacja polega na przeniesieniu się ze świata dwu- do świata trójwymiarowego. W przypadku binarnym wyróżniane były dwa stany: chory (ozn. 1), oraz zdrowy (ozn. 0). W przypadku testu różnicującego trzy grupy wprowadza się umownie oznaczenia  $D^+$  (zaawansowana choroba),  $D^0$  (pośrednie lub wczesne stadium),  $D^-$  (zdrowy). Wynik testu oznacza się przez  $T$ . Dla testów zakłada się, że wartości  $T$  rosną wraz z postępem choroby.

Podobnie jak poprzednio interesujące są: prawdopodobieństwo, że osoba o zaawansowanym stadium choroby zostanie przyporządkowana do odpowiedniej grupy —  $Pr(T > t_+ \mid D^+)$ , prawdopodobieństwo, że osoba zdrowa zostanie zaklasyfikowana do grupy zdrowych —  $Pr(T < t_- \mid D^-)$  oraz dodatkowo prawdopodobieństwo, że osoba z grupy  $D_0$  zostanie również prawidłowo sklasyfikowana —  $Pr(t_- < T < t_+ \mid D^0)$ , gdzie  $t_+$  oraz  $t_-$  (ustalone) oznaczają wartości graniczne wyników testu pomiędzy grupami, oraz zachodzi  $t_- < t_+$ .

Niech  $F_i$  oznacza ciągłą i ściśle rosnącą dystrybuantę funkcji testu dla grupy  $D^i$  dla  $i = +, 0, -$ ,  $G_i = 1 - F_i$  i niech

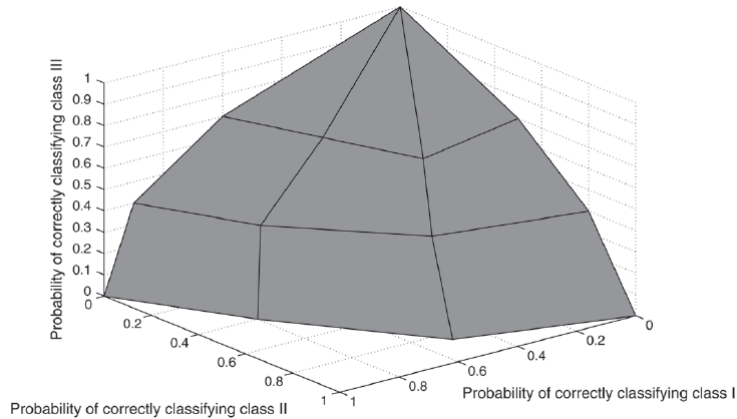
$$x = \Pr(T \leq t_- \mid D^-) = F_-(t_-), \quad (1.3)$$

$$y = \Pr(T \geq t_+ \mid D^+) = G_+(t_+), \quad (1.4)$$

$$z = z(x, y) = \Pr(t_- < T < t_+ \mid D^0) = F_0(t_+) - F_0(t_-) = F_0(G_+^{-1}(y)) - F_0(F_-^{-1}(x)), \quad (1.5)$$

gdzie  $F_-^{-1}$  i  $G_+^{-1}$  oznaczają funkcje odwrotne do  $F_-$  i  $G_+$  odpowiednio, natomiast  $z = z(x, y)$  definiuje powierzchnię w trójwymiarowej przestrzeni  $(x, y, z)$ , którą nazywa się przestrzenią ROC i stanowi rozszerzenie krzywej ROC na przypadek trójwymiarowy. Warunek  $t_- < t_+$  implikuje, że powierzchnia ROC jest poprawnie zdefiniowana w dziedzinie  $y < G_+(F_-^{-1}(x))$ . Analogicznie do przypadku dwóch wymiarów, wykres rozszerzonej ROC znajduje się w kostce jednostkowej pierwszej ćwiartki układu współrzędnych o trzech wymiarach o początku w punkcie  $(0, 0, 0)$ . Uwagę warto zwrócić na punkt  $(x, y, z) = (1, 1, 1)$  oznaczający, że dla danego testu diagnostycznego istnieje tak para punktów  $t_- < t_+$ , że test perfekcyjnie klasyfikuje wszystkich pacjentów do grupy  $D^+$ , jeśli jego wartość jest większa od  $t_+$ , do  $D^-$ , jeśli jest poniżej  $t_-$  i do grupy  $D^0$  dla wszystkich wartości pośrednich. Im bliżej powierzchnia ROC jest punktu „idealnego”, tym lepiej dopasowana para punktów oddzielających może być zastosowana do diagnozowania trzech grup.

**Rysunek 1.2:** Przykładowy wykres przedstawiający powierzchnię ROC w trójwymiarowej przestrzeni. Każda z osi układu odpowiada prawdopodobieństwu przypisania do odpowiedniej grupy. Układ wszystkich możliwych kombinacji wartości tych prawdopodobieństw dla danego markera stanowi powierzchnię ROC.



Źródło: Jialiang Li, „ROC analysis with multiple classes and multiple tests: methodology and its application in microarray studies.” [9].

Do oceny klasyfikatora rozróżniającego trzy grupy może posłużyć przestrzeń pod wykresem (ozn. VUS) wyznaczonym przez punkty  $(x, y, z)$ . W przypadku powierzchni ROC zauważana jest tendencja, że mniejsza odległość euklidesowa powierzchni ROC od punktu  $(1, 1, 1)$  wiąże się z większym prawdopodobieństwem, że para punktów  $t_-$  i  $t_+$  w dokładniejszy sposób rozdziela badanych do trzech grup.

Objętość ta, czyli VUS, obliczana jest ze wzoru (z [21]):

$$V_{00} = \iint_{D_{00}} F_0(G_+^{-1}(y)) - F_0(F_-^{-1}(x)) dy dx. \quad (1.6)$$

Dziedziną jest  $D_{00} = \{(x, y) | 0 \leq x \leq 1, 0 \leq y \leq G_+(F_-^{-1}(x))\}$ . Jeżeli test działa idealnie poprawnie, wartość  $V_{00}$  jest równa 1.

Objętość powierzchni pod rozszerzoną krzywą ROC reprezentuje prawdopodobieństwo, że dany test — wykorzystywany do niego marker, pozwoli poprawnie posortować trójkę elementów, zawierających losowo wybrany element z każdej z trzech populacji.

Dla testu, który zupełnie nie posiada zdolności poprawnej diagnozy, odpowiadająca powierzchnia ROC spełnia  $x + y + z = 1$ , czyli jest płaszczyzną w  $R^3$  zdefiniowaną jako  $\{(x, y) | x > 0, y > 0, x + y < 1\}$ . Wówczas  $V_{00} = \frac{1}{6}$ .

W przypadku testów diagnozujących choroby dla trzech grup niepożądanym jest jednocześnie niska wartość czułości i niska wartość swoistości dla grup  $D^+$  i  $D^-$ . Wartość VUS można również szacować przy sprecyzowanym założeniu o minimalnym poziomie tych dwóch wartości. Jeśli oznaczy się je odpowiednio przez  $q$  i  $p$  i założyć, że  $G_+^{-1}(q) \geq F_-^{-1}(p)$ , to częściową objętość pod powierzchnią ROC definiuje się następująco ([21]):

$$V_{pq} = \iint_{D_{pq}} [F_0(G_+^{-1}(y)) - F_0(F_-^{-1}(x))] dx dy, \quad (1.7)$$

gdzie  $D_{pq} = \{(x, y) | p \leq x \leq 1, q \leq y \leq G_+(F_-^{-1}(x))\}$ .

W przypadku perfekcyjnego testu  $V_{pq} = (1-p)(1-q)$ . Jeżeli  $x + y + z = 1$ , to w dziedzinie  $\{(x, y) | x > 0, y > 0, x + y < 1\}$

$$V_{pq} = \frac{(1-p)^3}{6} - \frac{q^3}{6} - \frac{(1-p)^2 q}{2} + \frac{(1-p)q^2}{2}. \quad (1.8)$$

Więcej na ten temat można znaleźć w artykule [21].

### 1.2.1. Estymacja VUS

W przypadku, gdy rozważa się trzy grupy oraz zakłada, że dana zmienna w każdej z nich ma rozkład normalny lub rozkład, który można przy pomocy transformacji monotonicznej przekształcić do normalnego, estymacja miary VUS może wykorzystywać metody parametryczne. W rozdziale tym zostanie przedstawiony sposób szacowania wartości VUS.

Niech  $N$  oznacza liczebność osób, których stan zdrowia zostaje przebadany. Na tej podstawie każdą z osób przyporządkowuje się do odpowiedniej grupy  $D^-$ ,  $D^+$  lub  $D^0$ . Niech  $n_i$  oznacza liczebność poszczególnych grup, tzn. liczbę wykonanych niezależnie testów,  $\mu_i$  – wartość oczekiwaną, a  $\sigma_i$  – odchylenie standardowe rozkładu danej zmiennej w poszczególnych grupach,  $i = +, 0, -$ , natomiast  $\hat{\mu}_i$  i  $\hat{\sigma}_i$  – ich nieobciążone estymatory największej wiarygodności.

Dla uproszczenia zapisu definiuje się także

$$a = \frac{\sigma_0}{\sigma_-}, \quad (1.9)$$

$$b = \frac{\mu_- - \mu_0}{\sigma_-}, \quad (1.10)$$

$$c = \frac{\sigma_0}{\sigma_+}, \quad (1.11)$$

$$d = \frac{\mu_+ - \mu_0}{\sigma_+}. \quad (1.12)$$

Odpowiednio przy pomocy estymatorów średniej z rozkładu i odchylenia standardowego wyznaczyć można estymatory  $\hat{a}, \hat{b}, \hat{c}, \hat{d}$ .

Wówczas powierzchnię ROC w układzie  $(x, y, z(x, y))$  wyznacza

$$z = \Phi \left[ \frac{\Phi^{-1}(1 - y) + \hat{d}}{\hat{c}} \right] - \Phi \left[ \frac{\Phi^{-1}(x) + \hat{b}}{\hat{a}} \right]. \quad (1.13)$$

Również szczegóły tej części można znaleźć w artykule [21].

### 1.3. Indeks Youdena

Indeks Youdena jest bardzo użyteczny przy określaniu własności markerów używanych w testach diagnostycznych. Odnosi się on bezpośrednio do czułości (def. 1.1.1) i swoistości (def. 1.1.2) i określa optymalny punkt przecięcia. Jest bardziej czytelny niż pole powierzchni pod krzywą ROC, ponieważ nie tylko obrazuje dyskriminującą dokładność testu diagnostycznego, ale również zapewnia najlepsze wartości sumy czułości i swoistości, gotowe do użycia w dalszych krokach.

Indeks ten dotyczy danych binarnych (dla dwóch grup) i jest definiowany jako

**Definicja 1.3.1** Indeks Youdena:  $J(t) = Se(t) + Sp(t) - 1$ ,

czyli stanowi kombinację czułości i swoistości w danym punkcie przecięcia  $t$ .

Indeks osiąga największą wartość (jeden), kiedy test daje dokładne rozróżnienie pomiędzy grupami zdrowych i chorych, a najmniejszą (wynoszącą zero), gdy przyporządkowuje z dokładnością nie lepszą niż przypadek.

Optymalny punkt przecięcia  $t^*$  jest osiągany dla takiego  $t$ , gdzie indeks osiąga maksymalną wartość, tzn.  $t^* = \arg \max_t J(t)$ . Graficznie, jest to największa odległość w linii prostej równoległej do osi OY pomiędzy krzywą ROC a diagonalą kwadratu jednostkowego pierwszej ćwiartki (która jest krzywą obrazującą przyporządkowanie do grup przez przypadek).

Obecne testy diagnostyczne odnoszą się najczęściej do dwóch grup, rozdzielających badanych na chorych i zdrowych. Problem trzech grup jest traktowany jako dwa rozważania dwóch grup, które prowadzą do znalezienia optymalnych punktów przecięcia. Luo i Xiong w [11] wprowadzają rozszerzony indeks Youdena, który odnosi się do trzech grup. Jest on zgeneralizowanym odpowiednikiem indeksu dla dwóch populacji.

Niech, podobnie jak poprzednio,  $D^+$ ,  $D^0$ ,  $D^-$  oznaczają odpowiednio: grupę w zaawansowanej fazie choroby, we wczesnej fazie choroby i zdrowych. Zakłada się również ciągłość wyników testu diagnostycznego ( $T$ ) w każdej z grup, a także jego monotoniczność — wraz z postępem choroby wyniki rosną (jeżeli monotonicznie maleją, to można wziąć wartości przeciwnie). Niech  $P_i$  będą gęstościami prawdopodobieństwa, że dany wynik zostanie poprawnie przypisany do odpowiedniej grupy, a  $F_i$  odpowiadającymi im ciągłymi i ściśle rosnącymi dystrybuantami dla grup  $D^i$ ,  $i = +, 0, -$ . Istnieje para progów dla testu diagnostycznego  $t_-$  i  $t_+$ , która pozwala zakwalifikować chorych do danej grupy. Zakłada się, że zachodzi nierówność  $t_- < t_+$ . Wtedy osoby z wynikami poniżej  $t_-$  trafiają do grupy  $D^-$ , ci z wynikami powyżej  $t_+$  do  $D^+$ , a pozostali do  $D^0$ .

Prawdopodobieństwa odpowiedniego przypisania do danej grupy definiuje się jako:

$$Sp(t_-) = P_-(T \leq t_-) = F_-(t_-), \quad (1.14)$$

$$Se(t_+) = P_+(T \geq t_+) = 1 - F_+(t_+), \quad (1.15)$$

$$Sm(t_-, t_+) = P_0(t_- \leq T \leq t_+) = F_0(t_+) - F_0(t_-), \quad (1.16)$$

kolejno: osoby zdrowej do  $D^-$ , chorej do  $D^+$ , we wczesnym stadium choroby do  $D^0$ .

Rozszerzony indeks Youdena, który ocenia diagnostyczną dokładność testu dla trzech grup, definiowany jest w [10] jako

$$J(t_-, t_+) = \frac{1}{2}(Sp(t_-) + Sm(t_-, t_+) + Se(t_+) - 1) = \quad (1.17)$$

$$= \frac{1}{2}(F_-(t_-) + F_0(t_+) - F_0(t_-) - F_+(t_+)). \quad (1.18)$$

Stanowi on więc kombinację prawdopodobieństw odpowiedniego przypisania wyniku danej osoby do odpowiedniej grupy; zależy od przyjętych punktów przejścia z jednej grupy do drugiej.

Optymalna para progów  $(t_-^*, t_+^*)$  jest wyznaczana poprzez maksymalizowanie indeksu po wszystkich możliwych parach. W takim punkcie stanowi on całkowitą dokładność testu. Indeks przyjmuje wartości od 0 (przyporządkowanie przez przypadek) do 1 (idealne rozróżnienie między grupami).

Istotne jest więc wyznaczenie optymalnych  $t_-^*$  i  $t_+^*$ , które maksymalizują indeks Youdena.

### 1.3.1. Estymacja parametryczna

Założmy, że rozkład obserwacji jest rozkładem normalnym. Wtedy  $P_i$  ze wzorów powyżej,  $i = +, 0, -$ , to gęstość rozkładu normalnego  $N(\mu_i, \sigma_i)$ , gdzie  $\mu_i$  to wartość oczekiwana, a  $\sigma_i$  to odchylenie standardowe. Bez straty ogólności zakładamy, że  $\mu_- < \mu_0 < \mu_+$ . Niech  $\Phi(x)$  oznacza dystrybuantę standardowego rozkładu normalnego. Mamy

$$Sp(t_-) = P_-(T \leq t_-) = \Phi\left(\frac{t_- - \mu_-}{\sigma_-}\right), \quad (1.19)$$

$$Se(t_+) = P_+(T \geq t_+) = \Phi\left(\frac{t_+ - \mu_+}{\sigma_+}\right), \quad (1.20)$$

$$Sm(t_-, t_+) = P_0(t_- \leq T \leq t_+) = \Phi\left(\frac{t_+ - \mu_0}{\sigma_0}\right) - \Phi\left(\frac{t_- - \mu_0}{\sigma_0}\right). \quad (1.21)$$

Zatem, przyjmując powyższe oznaczenia przy założeniu rozkładu normalnego, indeks Youdena można wyrazić jako

$$J(t_-, t_+) = \frac{1}{2} \left\{ \left[ \Phi\left(\frac{t_- - \mu_-}{\sigma_-}\right) - \Phi\left(\frac{t_- - \mu_0}{\sigma_0}\right) \right] + \left[ \Phi\left(\frac{t_+ - \mu_0}{\sigma_0}\right) - \Phi\left(\frac{t_+ - \mu_+}{\sigma_+}\right) \right] \right\}. \quad (1.22)$$

Optymalna para punktów granicznych jest otrzymywana poprzez wyliczenie pochodnych cząstkowych po  $t_-$  i  $t_+$  powyższego równania i następnie przyrównanie ich do zera. Można w ten sposób wrazić wzory na  $t_-^*$  i  $t_+^*$ :

$$t_-^* = \frac{(\mu_0 \sigma_-^2 - \mu_- \sigma_0^2) - \sigma_- \sigma_0 \sqrt{(\mu_- - \mu_0)^2 + (\sigma_-^2 - \sigma_0^2) \ln\left(\frac{\sigma_-^2}{\sigma_0^2}\right)}}{\sigma_-^2 - \sigma_0^2}, \quad (1.23)$$

$$t_+^* = \frac{(\mu_+ \sigma_0^2 - \mu_0 \sigma_+^2) - \sigma_+ \sigma_0 \sqrt{(\mu_0 - \mu_+)^2 + (\sigma_0^2 - \sigma_+^2) \ln(\frac{\sigma_0^2}{\sigma_+^2})}}{\sigma_0^2 - \sigma_+^2}. \quad (1.24)$$

Nierówność  $t_-^* < t_+^*$  nie zawsze musi zachodzić. W takim przypadku niemożliwe jest poprawne obliczenie wartości rozszerzonego indeksu Youdena jak i miary VUS, ponieważ dziedziny funkcji, które estymują te wskaźniki, są źle zdefiniowane. Ponadto  $t_-^*$  i  $t_+^*$  nie stanowią prawidłowych punktów progowych.

W przypadku, gdy równe są wariancje w grupach  $D^-$  i  $D^0$  lub w grupach  $D^0$  i  $D^+$ , punkty optimum przecięcia są wyznaczone tylko za pomocą wartości oczekiwanych:

$$t_-^* = \frac{\mu_- - \mu_0}{2}, \quad (1.25)$$

$$t_+^* = \frac{\mu_0 + \mu_+}{2}. \quad (1.26)$$

Kiedy założenie normalności nie jest uzasadnione, transformacja Boxa–Coxa jest często używana, żeby przybliżyć rozkład normalny. Transformacja ta przekształca monotonicznie dodatnio określoną zmienną  $x$  na zmienną  $y$  poprzez funkcję  $g$ :  $y = g(x, \lambda)$  o postaci

$$y = \frac{x^\lambda - 1}{\lambda} \quad (1.27)$$

dla  $\lambda \neq 0$ .

Dla  $\lambda \rightarrow 0$ , transformacja prowadzi do postaci logarytmicznej przekształcenia, to znaczy  $y = \log(x)$  dla  $\lambda = 0$ . W przypadku, gdy rozpatruje się trzy grupy, szacowanie  $\lambda$  może odbywać się dzięki numerycznym algorytmom optymalizacji (np. używając algorytmów z pakietu [16]).

### 1.3.2. Estymacja nieparametryczna

Kiedy żadne założenia dotyczące rozkładu nie mogą być poczynione, optymalny punkt przecięcia i rozszerzony indeks Youdena szacowane są metodami nieparametrycznymi. Dystrybuanty  $F_i$  dla  $i = +, 0, -$  mogą być otrzymywane albo empirycznie, albo za pomocą wygładzania jądrowego (*kernel smoothing*) na podstawie zaobserwowanych danych.

Niech  $n_i$ ,  $i = +, 0, -$  oznacza liczebność grupy  $D^i$ . Niech  $x_j^{(i)}$  oznacza wartość testu dla  $j$ -tej osoby w  $i$ -tej grupie. Empiryczne dystrybuanty tego testu dla trzech grup są estymowane jako  $\hat{F}_i(t) = \frac{1}{n_i} \sum_{j=1}^{n_i} I(x_j^{(i)} \leq t)$ ,  $i = +, 0, -$ , gdzie  $I(u \leq t)$  to indykator. Z wcześniej wybraną jądrową dystrybuantą  $K(x)$  i szerokością okna  $h$  (liczba dodatnia) można estymować funkcje gęstości jako

$$P_i(T = t) = \frac{1}{n_i} \sum_{j=1}^{n_i} \frac{1}{h_i} K\left(\frac{t - x_j^{(i)}}{h_i}\right), \quad (1.28)$$

$i = +, 0, -$ .

Wybór  $K(x)$  zazwyczaj nie jest istotny i często wybierane jest jądro Gaussowskie (*Gaussian kernel*). Szerokość  $h_i$  jest istotna do kontrolowania stopnia wygładzania. Warto podkreślić, że możliwy jest wybór innego  $h_i$  dla każdej z trzech grup. Zastosowanie wygładzania jądrowego Gaussa (*Gaussian kernel smoothing*) prowadzi do oszacowania dystrybuant rozkładu każdej z grup:

$$\hat{F}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \Phi\left(\frac{t - x_j^{(i)}}{h_i}\right), \quad i = +, 0, -.$$

Zastosowanie empirycznej metody pozwala oszacować indeks Youdena następująco:

$$\hat{J}(t_-, t_+) = \frac{1}{2} \left[ \frac{1}{n_-} \sum_{j=1}^{n_-} I(x_j^{(-)} \leq t_-) - \frac{1}{n_0} \sum_{j=1}^{n_0} I(x_j^{(0)} \leq t_-) + \right. \\ \left. + \frac{1}{n_0} \sum_{j=1}^{n_0} I(x_j^{(0)} \leq t_+) - \frac{1}{n_+} \sum_{j=1}^{n_+} I(x_j^{(+)} \leq t_+) \right]. \quad (1.29)$$

Natomiast metoda wygładzania jądrowego Gaussa dystrybucyjny prowadzi do następującego wzoru na indeks Youdena:

$$\hat{J}(t_-, t_+) = \frac{1}{2} \left[ \frac{1}{n_-} \sum_{j=1}^{n_-} \Phi \left( \frac{t_- - x_j^{(-)}}{h_-} \right) - \frac{1}{n_0} \sum_{j=1}^{n_0} \Phi \left( \frac{t_- - x_j^{(0)}}{h_0} \right) + \right. \\ \left. + \frac{1}{n_0} \sum_{j=1}^{n_0} \Phi \left( \frac{t_+ - x_j^{(0)}}{h_0} \right) - \frac{1}{n_+} \sum_{j=1}^{n_+} \Phi \left( \frac{t_+ - x_j^{(+)}}{h_+} \right) \right]. \quad (1.30)$$

Optymalna para progów  $(t_-^*, t_+^*)$  z powyższych równości nie może być otrzymana z analitycznych rozważań, konieczne jest użycie numerycznych algorytmów w celu otrzymania rozwiązania.

### 1.3.3. Parametryczne przedziały ufności

Przedział ufności określa, w jakim przedziale mieści się wartość szacowanego parametru. Ustalenie tego przedziału odbywa się z wcześniej określoną dokładnością — można sprecyzować, jak duże jest prawdopodobieństwo, że szacowana wielkość w nim wystąpi.

**Definicja 1.3.2** (z [15]) Niech  $g(\theta)$  będzie funkcją nieznanego parametru. Rozważmy dwie statystyki  $\underline{g} = \underline{g}(X)$  i  $\bar{g} = \bar{g}(X)$ .

Mówimy, że  $[\underline{g}, \bar{g}]$  jest przedziałem ufności dla  $g(\theta)$  na poziomie ufności  $1 - \alpha$ , jeśli

$$P_\theta \left( \underline{g}(X) \leq g(\theta) \leq \bar{g}(X) \right) \geq 1 - \alpha$$

dla każdego  $\theta$ .

Przy założeniu normalności rozkładu, zarówno optymalne punkty progowe, jak i indeks Youdena są funkcjami niezależnych od siebie estymatorów wartości oczekiwanych i odchyłeń standardowych  $(\hat{\mu}_-, \hat{\mu}_0, \hat{\mu}_+ \text{ i } \hat{\sigma}_-, \hat{\sigma}_0, \hat{\sigma}_+)$  z trzech niezależnych rozkładów normalnych, jednakże w tym przypadku przyjmowane jest w dalszym ciągu założenie o monotonicznym wzroście wartości oczekiwanych:  $\mu_- < \mu_0 < \mu_+$ . Stosując wieloczynnikową metodę delta (na podstawie [4]):

**Twierdzenie 1.3.1 (Wieloczynnikowa metoda delta)** Niech  $X^{(1)}, \dots, X^{(n)}$  będzie próbką losową o wartości oczekiwanej  $\mathbf{E}(X^{(i)}) = \mu$  i kowariancji  $\mathbf{E}(X^{(i)} - \mu)(X^{(i)} - \mu)^T = \Sigma$ . Dla danej funkcji  $g$ , która ma ciągłą pierwszą pochodną cząstkową, zachodzi

$$\sqrt{n} \left( g(\hat{X}) - g(\mu) \right) \rightarrow \mathcal{N} \left( 0, \nabla^T g(\mu) \Sigma \nabla g(\mu) \right) \text{ według rozkładu.}$$

Dowód w Dodatku A.

można wyliczyć asymptotyczne wariancje optymalnych punktów przecięcia oraz optymalny indeks Youdena korzystając z poniższych równań (z [11]):

$$\begin{aligned} Var(\hat{t}_-^*) &= \left(\frac{\partial t_-^*}{\partial \mu_-}\right)^2 Var(\hat{\mu}_-) + \left(\frac{\partial t_-^*}{\partial \mu_0}\right)^2 Var(\hat{\mu}_0) + \\ &+ \left(\frac{\partial t_-^*}{\partial \sigma_-}\right)^2 Var(\hat{\sigma}_-) + \left(\frac{\partial t_-^*}{\partial \sigma_0}\right)^2 Var(\hat{\sigma}_0), \end{aligned} \quad (1.31)$$

$$\begin{aligned} Var(\hat{t}_+^*) &= \left(\frac{\partial t_+^*}{\partial \mu_0}\right)^2 Var(\hat{\mu}_0) + \left(\frac{\partial t_+^*}{\partial \mu_+}\right)^2 Var(\hat{\mu}_+) + \\ &+ \left(\frac{\partial t_+^*}{\partial \sigma_-}\right)^2 Var(\hat{\sigma}_0) + \left(\frac{\partial t_+^*}{\partial \sigma_+}\right)^2 Var(\hat{\sigma}_+), \end{aligned} \quad (1.32)$$

$$\begin{aligned} Var(\hat{J}^*) &= \left(\frac{\partial J}{\partial \mu_-}\right)^2 Var(\hat{\mu}_-) + \left(\frac{\partial J}{\partial \mu_0}\right)^2 Var(\hat{\mu}_0) + \left(\frac{\partial J}{\partial \mu_+}\right)^2 Var(\hat{\mu}_+) + \\ &+ \left(\frac{\partial J}{\partial \sigma_-}\right)^2 Var(\hat{\sigma}_-) + \left(\frac{\partial J}{\partial \sigma_0}\right)^2 Var(\hat{\sigma}_0) + \left(\frac{\partial J}{\partial \sigma_+}\right)^2 Var(\hat{\sigma}_+). \end{aligned} \quad (1.33)$$

Estymatory punktów przecięcia i indeksu Youdena są asymptotycznie nieobciążone (obciążenie estymatora dąży do zera przy rosnącej liczbie próby) i mają rozkład normalny, zatem przedział ufności na poziomie istotności  $\alpha \times 100\%$  może być otrzymany jako  $\hat{t}_-^* \pm z_{\alpha/2} \sqrt{Var(\hat{t}_-^*)}$ ,  $\hat{t}_+^* \pm z_{\alpha/2} \sqrt{Var(\hat{t}_+^*)}$ ,  $\hat{J}^* \pm z_{\alpha/2} \sqrt{Var(\hat{J}^*)}$ , gdzie  $z_{\alpha/2}$  reprezentuje  $\alpha/2$  kwantyl standardowego rozkładu normalnego (literatura: [11]). Zatem za pomocą obliczonych wzorów można otrzymać przedział ufności dla rozszerzonego indeksu Youdena na poziomie ufności  $(1 - \alpha) \times 100\%$ . Jednakże,  $t_-^*$  i  $t_+^*$  nie są już niezależne od siebie, ponieważ oba zależą od  $\mu_0$  i  $\sigma_0$ . Kowariancja dwóch estymatorów jest otrzymywana z rozwinięcia Taylora:

$$Cov(\hat{t}_-^*, \hat{t}_+^*) = \frac{\partial t_-^*}{\partial \mu_-^*} \frac{\partial t_+^*}{\partial \mu_0^*} Var(\hat{\mu}_0) + \frac{\partial t_-^*}{\partial \sigma_-^*} \frac{\partial t_+^*}{\partial \sigma_0^*} Var(\hat{\sigma}_0). \quad (1.34)$$

Wariancje estymatorów próbkowych z parametrami rozkładu normalnego są znane:

$$Var(\hat{\mu}_i) = \frac{1}{n_i} \sigma_i^2, \quad Var(\hat{\sigma}_i) = \frac{1}{2(n_i - 1)} \sigma_i^2, \quad i = +, 0, -. \quad (1.35)$$

Podstawiając wzory 1.35 na wariancje i pochodne cząstkowe pojawiające się we wzorach 1.31, 1.32, 1.33 (policzone w Dodatku B) otrzymujemy wzory na asymptotyczną wariancję i kowariancję wyliczonych optymalnych punktów przecięcia i wariancję oszacowanej wartości optymalnego indeksu Youdena.

#### 1.3.4. Nieparametryczne przedziały ufności

Gotowa formuła na obliczenie optymalnych punktów przecięcia oraz indeksu Youdena przy braku założenia normalności rozkładu nie jest znana. Jednakże, podstawowe kwantyle przedziału ufności metody bootstrap mogą zostać obliczone. Tworzone są próbki bootstrapowe, z



każdej otrzymywane są estymatory  $t_-^*$ ,  $t_+^*$  i  $J^*$  i następnie wyliczone zostają kwantyle dolnej i górnej granicy tych estymatorów, odpowiednio  $\alpha/2$  i  $1 - \alpha/2$ .

Metoda bootstrap może być również zastosowana w przypadku, gdy rozkład zmiennej jest znany. Procedura znajdowania wartości granicznych przedziałów ufności wygląda tak samo, aczkolwiek ze względu na jej wysoki koszt obliczeniowy, często jest nieopłacalna.



## Rozdział 2

# Funkcje pakietu R do analizy klasyfikatorów: biblioteka `DiagTest3Grp`

Ocena klasyfikatorów wymaga przeprowadzenia znacznej ilości obliczeń na danych, których rozkład dla określonej próby nie jest znany. Pomocnym okazuje się użycie pakietu R. W 2010 r. Jingqin Luo oraz Chengjie Xiong stworzyli bibliotekę dla tego pakietu, która umożliwia przeprowadzenie statystycznej analizy testów diagnostycznych dla trzech grup, które zostały opisane w poprzednim rozdziale. Biblioteka `DiagTest3Grp`, o której mowa, jest dostępna dla wszystkich zainteresowanych online [12], podobnie jak jej pełna dokumentacja ([13]).

W rozdziale tym zostaną dokładniej przedstawione jedynie te funkcje, które zostały wykorzystane do przeprowadzenia analizy w tej pracy.

Dostępny pakiet `DiagTest3Grp` jest wyposażony w bazę danych dotyczącą choroby Alzheimera. Jest to ten sam zestaw, którego analiza jest dość dokładnie opisana w pracy [10]. Również w dostępnej pomocy dotyczącej tego pakietu można znaleźć przykłady zastosowania funkcji na tych danych. W celu uniknięcia pomyłek dotyczących oznaczeń w rozdziale tym w poleceniach zostaną wykorzystywane ogólne oznaczenia argumentów funkcji. W następnych rozdziałach pracy zaprezentowane zostanie zastosowanie `DiagTest3Grp` na danych rzeczywistych dotyczących zaburzeń psychiatrycznych.

Do analizy klasyfikatorów konieczna jest baza danych, której zmiennymi są:

- klasyfikator, który dzieli obserwacje na trzy grupy — są to „punkty odniesienia”, które warunkują prawdopodobieństwa poprawnej klasyfikacji pozostałych markerów,
- inne klasyfikatory, których statystyczna poprawność w rozróżnianiu grup, istotność oraz wzajemne wzmacnianie zostanie poddana analizie.

W niemalże wszystkich poniższych funkcjach stosowane jest założenie, że wartości markerów rosną monotonicznie wraz z rozwojem choroby.

### 2.1. Funkcje `Youden3Grp` i `VUS`

Do wyestymowania na podstawie danych wartości objętości pod wykresem ROC używa się funkcji `VUS`. Kod wywołujący funkcję `VUS` jest postaci:

```
VUS(x, y, z, method = c("Normal", "NonPar"), p = 0, q = 0, alpha = 0.05,  
NB00T = 100, subdivisions = 50000, lam.minus = 1/3, lam0 = 1/3, lam.plus
```

```
= 1/3, typeIerror = 0.05, margin =0.05,FisherZ=FALSE,optimalCut=TRUE,
cut.seq=NULL,optimize=FALSE)
```

W Tabeli 2.1 opisany został zbiór najważniejszych argumentów tej funkcji, których odpowiednie dobranie ma kluczowe znaczenie dla przeprowadzanych obliczeń.

Ogólny schemat komendy dla przykładowych danych ma postać:

```
x <- kfront[group=="D-"]
y <- kfront[group=="D0"]
z <- kfront[group=="D+"]

##normal estimate
normal.res <- VUS(x,y,z,method="Normal",p=0,q=0,alpha=0.05)
normal.res
```

Funkcja ta zwraca niemalże wszystkie informacje, które są interesujące podczas diagnostyki testu. Po pierwsze jest to wartość miary VUS dla danego klasyfikatora oraz jego przedział ufności (ozn. CI) dla zadanego poziomu istotności. Następnie zwracana jest także para punktów  $(t_-, t_+)$  określająca, jakie wartości danego markera najlepiej wybrać do rozdzielania trzech grup tak, aby zmaksymalizować szanse na poprawną klasyfikację niezależnie od doboru pacjenta. Stąd wynikiem są także współrzędne punktu  $(Sp, Sm, Se)$ , gdzie dla przypomnienia  $Sp = Pr(x \leq t_-)$  dla  $x \in D^-$ ,  $Se = Pr(z \geq t_+)$  dla  $z \in D^-$  oraz  $Sm = Pr(t_- \leq y \leq t_+)$  dla  $y \in D^0$  dane jako najmniejszy kwadrat odległości od punktu  $(1, 1, 1)$ , o którym powiedziane zostało już w podrozdziale 1.2. Dana funkcja zwraca ponadto liczebność próbki potrzebną do oszacowania VUS w określonych granicach błędu. Jest to rzecz bowiem niezwykle istotna w przypadku potwierdzenia diagnostycznej „zdolności” danego markera w realnych badaniach. W wyniku wyświetlane są również podstawowe statystyki dla danego klasyfikatora: liczba obserwacji w próbie, wartość średnia oraz odchylenie standardowe dla każdej z podgrup.

**Tabela 2.1:** Opis argumentów funkcji VUS

VUS	
Argument	Opis
$x, y, z$	wektor wyników testu diagnostycznego dla odpowiednio: $D^-$ , $D^0$ , $D^+$ ;
$method$	charakterystyka argumentu – wybór metody testu: „Normal” w przypadku spełnionego założenia o rozkładzie normalnym wyników lub „NonPar” w przypadku nie spełnienia tego założenia;
$p, q$	żądane minimalne wartości dla odpowiednio swoistości i czułości, $p, q \in (0, 1]$ ;
$alpha$	parametr określający poziom istotności;
$NBOOT$	liczba prób bootstrap – konieczne podanie w przypadku estymacji nieparametrycznej.

W przypadku, gdy dla każdej wyróżnionej grupy zmienna ma rozkład normalny lub do niego zbliżony, narzędziem jest estymacja parametryczna omówiona w podrozdziale 1.2. W przypadku danych medycznych bardzo często zdarza się, że rozkład ten nie przypomina rozkładu normalnego i nie można go przy pomocy transformacji monotonicznej do niego przekształcić lub jest nieznany. Problem pojawia się także w przypadku bardzo małych prób. Wówczas konieczna jest estymacja nieparametryczna, którą umożliwia pakiet *DiagTest3Grp*,

wykorzystującą dystrybuantę empiryczną, która liczy odsetek poprawnych porządków wśród wszystkich możliwych trójek, dla których każdy element pochodzi z innej grupy. Więcej o tej metodzie w podrozdziale 1.2. W celu jak najlepszego oszacowania wykorzystuje się metodę bootstrap.

Kolejnym wykorzystywanym poleceniem w pakiecie R jest funkcja `Youden3Grp` służąca estymacji na podstawie danych rozszerzonego indeksu Youdena. Kod wywołujący:

```
Youden3Grp(x, y, z, method = c("Normal", "TN", "EMP", "KS", "KS-SJ"),
randomStart.N = 1, optim.method = NULL, t.minus.start = NULL,
t.plus.start = NULL, lam.minus = 1/3, lam0 = 1/3, lam.plus = 1/3,
typeIerror = 0.05, margin=0.05, NBOOT=10, seed.seq=NULL, alpha=0.05,
FisherZ=FALSE)
```

Opis najważniejszych argumentów dla tej funkcji znajduje się w Tabeli 2.2.

**Tabela 2.2:** Opis argumentów funkcji `Youden3Grp`

Youden3Grp	
Argument	Opis
$x, y, z$	wektor wyników testu diagnostycznego dla odpowiednio: $D^-$ , $D^0$ , $D^+$ ;
<i>method</i>	charakterystyka argumentu — wybór metody testu: „Normal” w przypadku spełnionego założenia o rozkładzie normalnym wyników, „TN”- następuje transformacja Boxa-Coxa, a następnie estymacja przy założeniu o rozkładzie normalnym, „EMP”- estymacja indeksu poprzez wykorzystanie dystrybuanty empirycznej rozkładu, „KS”- zastosowanie estymatora jądrowego gęstości;
<i>alpha</i>	parametr określający poziom istotności;
<i>NBOOT</i>	liczba prób bootstrap — konieczne podanie w przypadku estymacji nieparametrycznej;
<i>margin</i>	granica błędu estymacji indeksu Youdena w obliczeniach wielkości próby.

Dla zdefiniowanych jak wyżej  $x, y, z$ , przykładowa komenda wygląda następująco:

```
youden <- Youden3Grp(x=x, y=y, z=z, method = "Normal")
youden
```

Podobnie jak dla funkcji `VUS`, rozszerzony indeks Youdena może być estymowany metodami parametrycznymi i nieparametrycznymi. Wszystkie wykorzystywane — zawarte w Tabeli 2.2 zostały szczegółowo opisane w podrozdziale 1.3.

Także i dla tej miary po wywołaniu uzyskiwane są statystyki dla trzech grup: liczebność, średnia i odchylenie standardowe, wyestymowane na podstawie próby rozszerzony indeks Youdena oraz przedział ufności dla jego wartości przy zadanym poziomie istotności, optymalny punkt rozdzielający ( $t_-$ ,  $t_+$ ), liczebność próbki konieczna do oszacowania wartości indeksu przy zachowaniu marginesu błędu oraz punkt ( $Sp$ ,  $Sm$ ,  $Se$ ), którego interpretacja się nie zmienia.

W celu graficznego przedstawienia uzyskanych wyników można dla każdej z powyższych miar wykonać komendę:

```
par(pty="m")
plot("vus", "youden")
```

Ostatnie polecenie w zależności od określonego argumentu rysuje wykres rozproszenia analizowanego klasyfikatora z zaznaczonymi punktami ( $t_-$ ,  $t_+$ ) oraz wykres pudełkowy. Oba wykresy wspomagają proces analizy wykorzystywanych markerów oraz ułatwiają interpretację dostępnych danych. Dla obu, dane pochodzące z grup  $D^-$ ,  $D^0$ ,  $D^+$  zaznaczane są odpowiednio kolorami: zielonym, niebieskim i czerwonym.

## 2.2. Inne funkcje pakietu DiagTest3Grp

W przypadku występowania kilku klasyfikatorów, które być może mogą być wykorzystane w diagnostyce choroby, rodzi się pytanie, który klasyfikator wybrać spośród dostępnych, jeżeli ich podstawowe miary są zbliżone lub co gorsza nie można ich porównać, lub jest to trudne. Do wspólnej analizy markerów służy polecenie `DiagTest3Grp.Test` oraz polecenie `Pairwise.DiagTest3Grp.Test`.

`DiagTest3Grp.Test` to test statystyczny do porównywania oszacowanych miar (VUS oraz Youden) dla dwóch lub więcej klasyfikatorów. W dostępnym pakiecie analiza wymaga założenia o rozkładzie normalnym dla każdej z grup diagnostycznych.

Opis argumentów znajduje się w Tabeli 2.3. Kod wywołujący jest postaci:

```
DiagTest3Grp.Test(dat,paired=FALSE,type=c("VUS","Youden"),p=0,q=0,mu=0,
conf.level=0.95,alternative=c("two.sided","less","greater")).
```

Niech  $\theta_j$  oznacza wartość miary VUS lub Youdena dla  $j$ -tego klasyfikatora.

W przypadku porównywania dwóch markerów test jest wykonywany przy hipotezie zerowej  $H_0 : \theta_1 - \theta_2 = \theta_0$ , która zakłada, że nieznaną wartość w populacji  $\theta$  jest równa wartości hipotetycznej  $\theta_0$ , przeciwko hipotezie alternatywnej  $H_1 : \theta_1 - \theta_2 \neq 0$  dla dwustronnej alternatywy:  $H_1 : \theta_1 - \theta_2 > 0$  dla *alternatywa* = „greater” ozn. *wiekszy* lub  $H_1 : \theta_1 - \theta_2 < 0$  dla *alternatywa* = „less” ozn. *mniej*. Ponieważ spełnione jest założenie o rozkładzie normalnym, do testowania hipotezy zerowej wykorzystywany jest  $Z$ -test, dla którego statystyka jest postaci  $z = \frac{\theta_1 - \theta_2}{SE(\theta_1 - \theta_2)}$ . Dla hipotezy zerowej statystyka ta ma asymptotyczny rozkład normalny ze średnią 0 i wariancją 1. Dla zadanego  $\alpha$  hipoteza zerowa zostaje odrzucona, jeśli  $|z| > z_{\alpha/2}$ , gdzie  $z_{\alpha/2}$  jest  $100 \times (1 - \frac{\alpha}{2})$  percentylem standardowego rozkładu normalnego. Wtedy również można utworzyć  $(1 - \alpha) \times 100\%$  przedział ufności dla

$$\theta_i - \theta_j = \hat{\theta}_i - \hat{\theta}_j \pm z_{\alpha/2} \sqrt{\hat{\sigma}_{\theta_i}^2 - \hat{\sigma}_{\theta_j}^2} \quad (2.1)$$

[21].

Wynikiem tej procedury jest wartość statystyki  $Z$ ,  $p$ -value,  $100 \times (1 - \alpha)\%$  przedział ufności oraz wartość estymacji miar dla danej pary.

W sytuacji porównywania  $n \geq 3$  markerów zbiorowy test, podobnie jak jednoczynnikowa analiza wariancji, jest przeprowadzany przy hipotezie zerowej  $H_0 : \theta_1 = \theta_2 = \dots = \theta_n$  przeciwko hipotezie alternatywnej  $H_1 : \exists i, j$  takie, że  $\theta_i \neq \theta_j$ . Statystyka testowa dąży asymptotycznie do statystyki chi-kwadrat z  $n - 1$  stopniami swobody ( $\chi^2(n - 1)$ ). Podobnie odrzucenie hipotezy zerowej następuje w przypadku, gdy  $\chi^2 > \chi_{\alpha}^2(n - 1)$  dla ustalonej  $\alpha$ , a  $\chi_{\alpha}^2$  jest  $\alpha \times 100\%$

percentylem  $\chi^2(n-1)$ .

**Tabela 2.3:** Opis argumentów funkcji `DiagTest3Grp.Test`

DiagTest3Grp.Test	
Argument	Opis
<i>dat</i>	wektor wyników testu diagnostycznego dla odpowiednio: $D^-$ , $D^0$ , $D^+$ ;
<i>paired</i>	wartość logiczna TRUE/FALSE wskazująca, czy wszystkie klasyfikatory są porównywane w tym samym zestawie;
<i>type</i>	określa miarę, według której rozpatrywana zostaje charakterystyka klasyfikatorów. Przyjmuje wartość „VUS” lub „Youden”;
<i>p</i> , <i>q</i>	parametry określające dolny poziom odpowiednio swoistości i czułości;
<i>mu</i>	hipotetycznie realna wartość różnicy pomiędzy wartościami danej miary pomiędzy markerami. Domyślnie jest równa 0;
<i>conf.level</i>	przedział ufności dla wartości bezwzględnej różnicy miar dwóch markerów. W przypadku większej liczby klasyfikatorów przyjmuje wartość NA;
<i>alternative</i>	określa rodzaj hipotezy alternatywnej, przyjmuje jedną z wartości: „two.sided”, „less”, „greater”.

Kolejnym poleceniem jest `Pairwise.DiagTest3Grp.Test`. Służy on do porównywania wszystkich par klasyfikatorów w celu sprawdzenia, które z nich charakteryzują się odrębną zdolnością diagnostyczną.

Schemat polecenia jest następujący:

```
Pairwise.DiagTest3Grp.Test(dat,paired=FALSE,type=c("VUS","Youden"),
p=0,q=0,mu=0,conf.level=0.95,alternative=c("two.sided","less","greater"),
p.adjust.method=c("none","holm","hochberg","hommel","bonferroni","BH","BY",
"fdr"),digits=3)
print(pairwise.youdenDolarprint.matrix[1:15, 1:7], na = "")
```

Jak widać, jego składowe są niemal identyczne z funkcją `DiagTest3Grp.Test`, dlatego też tabelą argumentów może być Tabela 2.3, do której należy jedynie dodać czynnik, który określa sposób obliczania *p-value* dopasowanego.

Wynikiem tej procedury jest macierz górnotrójkątna, w której dla każdej pary klasyfikatorów znajduje się wartość statystyki *z*, *p-value* oraz *p-value dopasowane*, które także dla lepszej interpretacji wyświetlone zostają na mapie ciepła. Postać tabeli oraz grafiki została dla konkretnych danych zaprezentowana w podrozdziale 3.4.





## Rozdział 3

# Analiza danych medycznych

W tym rozdziale przedstawione zostanie wykorzystanie pakietu `DiagTest3Grp` na danych rzeczywistych dotyczących chorób psychiatrycznych.

### 3.1. Opis zbioru danych

Analizowany w tej pracy zbiór danych to element pakietu `PBImisc`, który można znaleźć na stronie `CRAN.R Project` ([3]). Bazuje on na rzeczywistych danych medycznych (jest podzbiorem z losowymi modyfikacjami) dotyczących oceny dziennej hospitalizacji w Europie (*Euro-pean day hospital evaluation*). Obserwacjami są wyniki badań i informacje o pacjentach klinik psychiatrycznych. Zmienne tego zbioru danych są przedstawione w Tabeli 3.1.

**Tabela 3.1:** Zmienne zbioru danych EDEN

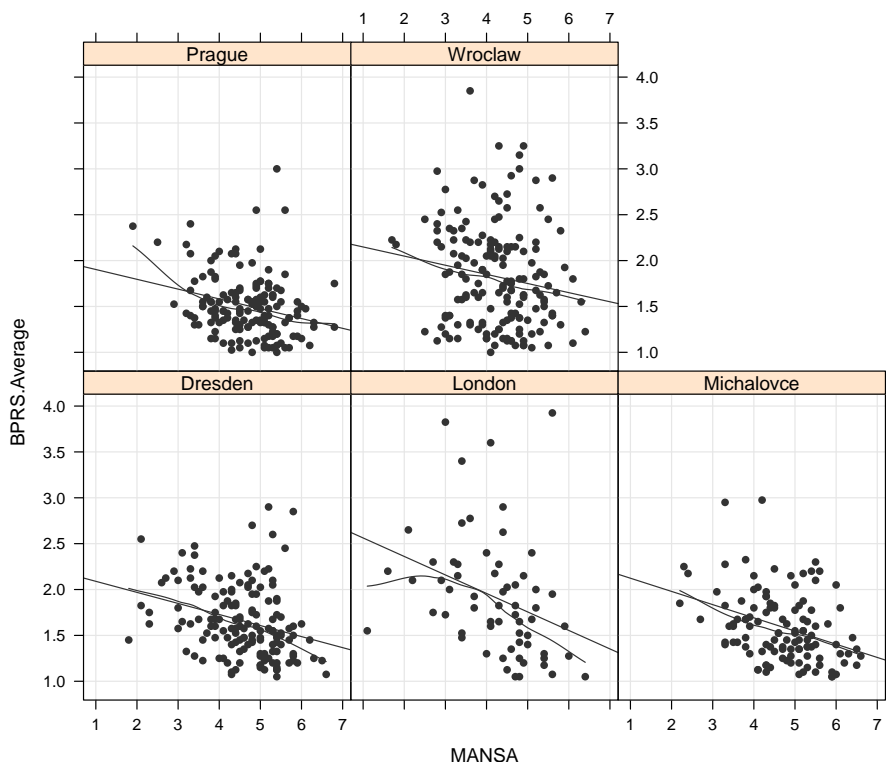
Zmienna	Opis
<code>mhid</code>	identyfikator lekarza, który badał danego pacjenta;
<code>center</code>	miasto, w którym badanie miało miejsce: Drezno, Londyn, Michalovce, Praga, Wrocław;
<code>BPRS.Maniac</code>	kolejne wyniki badania BPRS;
<code>BPRS.Negative</code>	
<code>BPRS.Positive</code>	
<code>BPRS.Depression</code>	
<code>BPRS.Average</code>	średnia z 24 pytań;
<code>MANSA</code>	jakość życia;
<code>sex</code>	płeć;
<code>children</code>	liczba posiadanych dzieci;
<code>years.of.education</code>	liczba lat spędzonych na edukacji;
<code>day</code>	tryb hospitalizacji: dzienny lub stacjonarny.

Objawy psychopatologiczne pacjentów szacowane są za pomocą BPRS, czyli *The Brief Psychiatric Rating Scale* ([1]). Jest to skala (od 1 – „nieobecny” do 7 – „ekstremalnie nasilony”) umożliwiająca ocenę postępu tych objawów wśród pacjentów psychiatrycznych. Składa się z 24 punktów odpowiadających poszczególnym symptomom. *Maniac* stanowi wyniki części badania odpowiedzialnej za zachowania maniakalne, a *Depression* – za tendencje do depresji. Przejawy pozytywne (*Positive*) odnoszą się do zakłóceń funkcji niezbędnych do normalnego

funkcjonowania (np. omamy i urojenia), natomiast negatywne (*Negative*) – do zacierania się lub ich utraty.

W tej pracy szczególna uwaga zwrócona jest na zmienną *MANSA*. Skrót ten pochodzi od *Manchester Short Assessment of Quality of Life*. Stanowi subiektywną ocenę jakości życia. Opiera się głównie na oszacowaniu wartości ogólnej satysfakcji z życia oraz poszczególnych jego obszarów (np. zatrudnienie, relacje rodzinne, hobby). Traktuje się ją jako główny klasyfikator i na jej podstawie dokonywany jest podział na trzy grupy. Rysunek 3.1 przedstawia wykres rozrzutu zmiennych *MANSA* i *BPRS.Average* w zależności od ośrodka, w którym wykonane zostało badanie.

**Rysunek 3.1:** Zależność zmiennych *MANSA* i *BPRS.Average* od miasta badania



W celu uzyskania trzech grup, obserwacje zostały podzielone ze względu na wartość zmiennej *MANSA*. Pacjenci z wynikiem mniejszym niż 4,2 trafiają do grupy  $D^-$ , z wynikiem większym bądź równym 5 – do grupy  $D^+$ , pozostali – do  $D^0$ . Przedziały te zostały dobrane arbitralnie tak, żeby liczebności grup były jak najbardziej do siebie zbliżone.

W analizie pod uwagę nie będą brane zmienne: *mdid*, *center*, *sex*, *children*, *years.of.education* oraz *day*, ponieważ nie oczekuje się po nich dyskryminującego wyniku. Pozostałe zmienne, czyli te związane ze skalą *BPRS*, są w dalszej części pracy poddawane testom w celu sprawdzenia ich własności.

Wczytany zbiór danych ([3]) zawiera 642 obserwacje. Dodana została zmienna dyskretna *group* przyjmująca wartości  $D^-$ ,  $D^0$ ,  $D^+$  w zależności od wartości zmiennej *MANSA*.

Liczebność każdej z grup przedstawia się następująco:  $D^-$  – 211,  $D^0$  – 209,  $D^+$  – 222.

Założeniem koniecznym do przeprowadzania testów w pakiecie *DiagTest3Grp* jest monotoniczność średnich w każdej grupie. Ścisłej, średnie wartości poszczególnych zmiennych mają rosnąć od  $D^-$  do  $D^+$ . W przypadku zbioru danych *EDEN* otrzymywana jest własność

przeciwna, czyli średnie wartości maleją wraz z postępem choroby. Dlatego też wszystkie wartości obserwacji interesujących zmiennych brane są ze znakiem przeciwnym i w ten sposób osiągnięta jest wymagana monotoniczność. Poszczególne wartości średnie przedstawione są w Tabeli 3.2.

**Tabela 3.2:** Średnie w każdej z grup dla kolejnych zmiennych

	<i>MANSA</i>	<i>Maniac</i>	<i>Negative</i>	<i>Positive</i>	<i>Depression</i>	<i>Average</i>
$D^-$	3,44	-1,41	-1,72	-1,52	-2,68	-1,83
$D^0$	4,56	-1,36	-1,67	-1,46	-2,18	-1,67
$D^+$	5,47	-1,30	-1,57	-1,28	-2,00	-1,54

## 3.2. Badanie rozkładu

W celu jak najlepszego przeanalizowania danych markerów trzeba poznać ich rozkład. Jest to ważne w świetle odpowiedniego szacowania wskaźników takich jak indeks Youdena czy VUS, a także sprawdzenia spełnienia założeń niektórych funkcji.

Podstawowym narzędziem, które umożliwia poznanie rozkładu danych empirycznych, jest histogram danej zmiennej. Histogramy badanych w tej pracy markerów są przedstawione na wykresie 3.2. Jak widać, rozpatrując każdą ze zmiennych, rozkład nie jest zbliżony do rozkładu normalnego.

### 3.2.1. Test Shapiro–Wilka

W celu potwierdzenia braku rozkładu normalnego, przeprowadzany jest test Shapiro – Wilka ([17]). Pozwala on wysnuć hipotezę, czy dana zmienna charakteryzuje się rozkładem normalnym. Test ten jest często stosowany w badaniach statystycznych.

Jest to test oparty o statystyki pozycyjne próby. Obserwowane są niezależne realizacje zmiennej losowej  $x_i$ ,  $i = 1, \dots, n$  o rozkładzie  $F$ . Hipotezą zerową jest rozkład normalny tej zmiennej o nieznannej wartości oczekiwanej  $\mu$  i nieznannej wariancji  $\sigma^2$ , czyli

$$H_0 : \exists_{\mu, \sigma} F = \mathcal{N}(\mu, \sigma^2)$$

przeciwko hipotezie

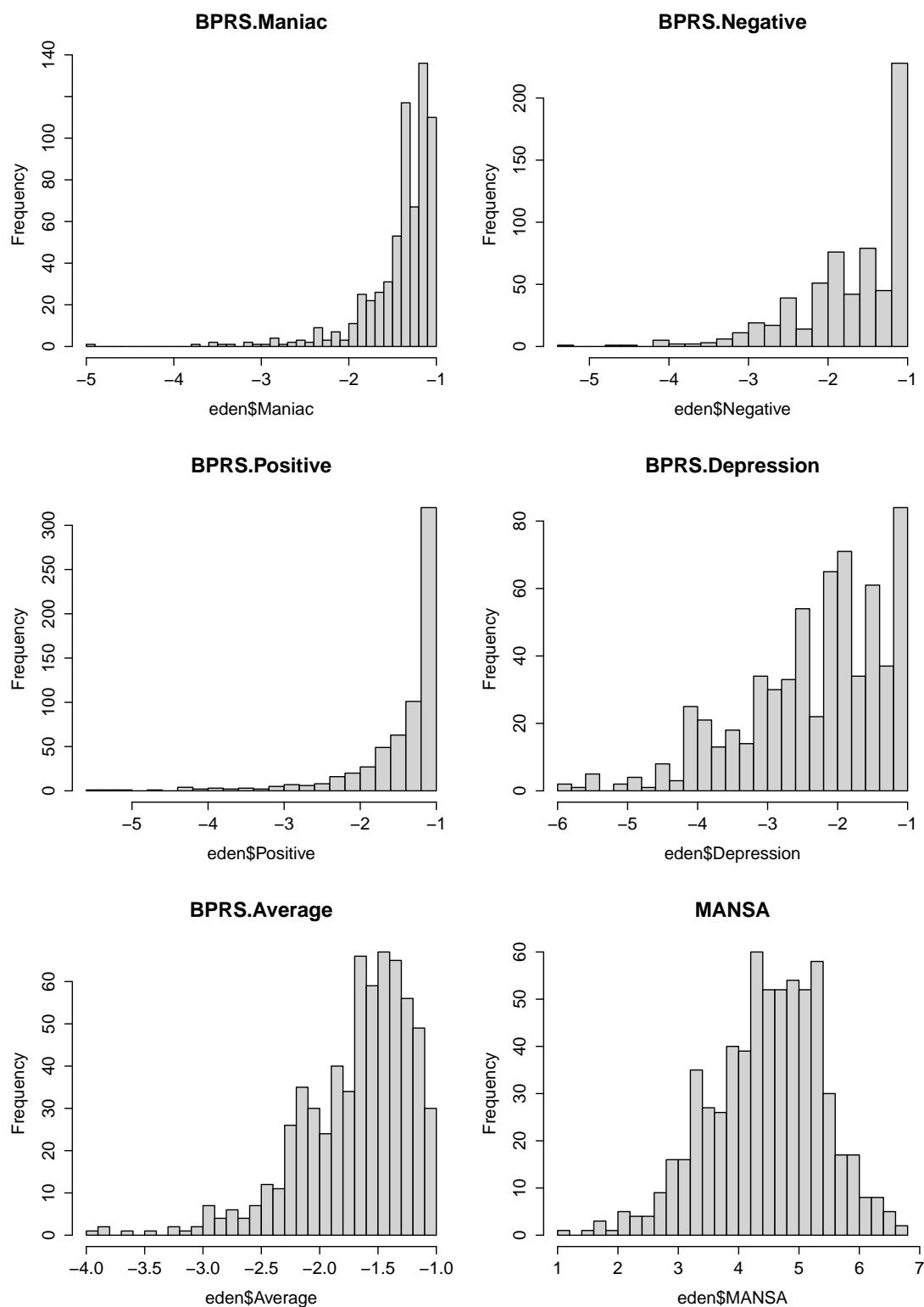
$$H_1 : \neg \exists_{\mu, \sigma} F = \mathcal{N}(\mu, \sigma^2).$$

Niech  $\vec{m}^T = (m_1, m_2, \dots, m_n)$  będzie wektorem wartości oczekiwanych uporządkowanych statystyk pozycyjnych z rozkładu normalnego i niech  $V = (v_{ij})$ ,  $i, j = 1, \dots, n$  będzie odpowiadającą macierzą kowariancji wymiaru  $n \times n$ . Wektor  $\vec{y}^T = (y_1, y_2, \dots, y_n)$  to wektor uporządkowanych statystyk pozycyjnych realizacji zmiennych losowych  $x_i$ , tj.  $y_i = x_{i:n}$ , gdzie  $x_{1:n} \leq x_{2:n} \leq \dots \leq x_{n:n}$ .

Żeby otrzymać wartość statystyki testowej  $W$  mając daną wielkość próbki  $n$  oraz wartości obserwacji  $x_1, x_2, \dots, x_n$ , należy wykonać następujące kroki:

- (i) uszeregować obserwacje  $x_i$ , żeby otrzymać statystykę pozycyjną próbki  $y_1 \leq y_2 \leq \dots \leq y_n$ ,

**Rysunek 3.2:** Histogramy badanych markerów.



(ii) obliczyć wartość  $S^2 = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (x_i - \bar{x})^2$ ,

(iii) a) jeżeli  $n$  jest liczbą parzystą, tzn.  $n=2k$ , to  $b$  jest dane wzorem:

$$b = \sum_{i=1}^k a_{n-i+1}(y_{n-i+1} - y_i),$$

gdzie wartości  $a_{n-i+1}$  można znaleźć w tablicach,

b) jeżeli  $n=2k+1$ , to obliczenia są analogiczne do a), z tym, że  $a_{k+1} = 0$  dla nieparzystego  $n$ . Z tego wynika, że  $b$  wyrażone jest wzorem:

$$b = a_n(y_n - y_1) + \dots + a_{k+2}(y_{k+2} - y_k),$$

(iv) obliczyć wartość  $W = \frac{b^2}{S^2}$ .

Dla danych obserwacji, oprócz wyliczonej wartości statystyki  $W$ , można wyznaczyć  $p$ -value przy określonym poziomie istotności (w tej pracy przyjęto poziom istotności równy 5%).

Sprawdzenie, czy dana zmienna ma rozkład normalny, może zostać przedstawione za pomocą wykresu kwantylowego ( $Q$ - $Q$  plot).

Na osiach wykresu  $Q$ - $Q$  znajdują się kwantyle badanej próbki i próbki o zakładanym, sprawdzanym rozkładzie (w tym przypadku o rozkładzie normalnym). Jeżeli dane obserwacje mają zakładany rozkład, to kolejne punkty leżą na jednej prostej (wychodzącej z punktu (0,0) pod kątem  $45^\circ$ ).

Przeprowadzone zostały testy Shapiro dla rozważanego zbioru danych za pomocą następującej funkcji (przykład dla markera BPRS.Maniac):

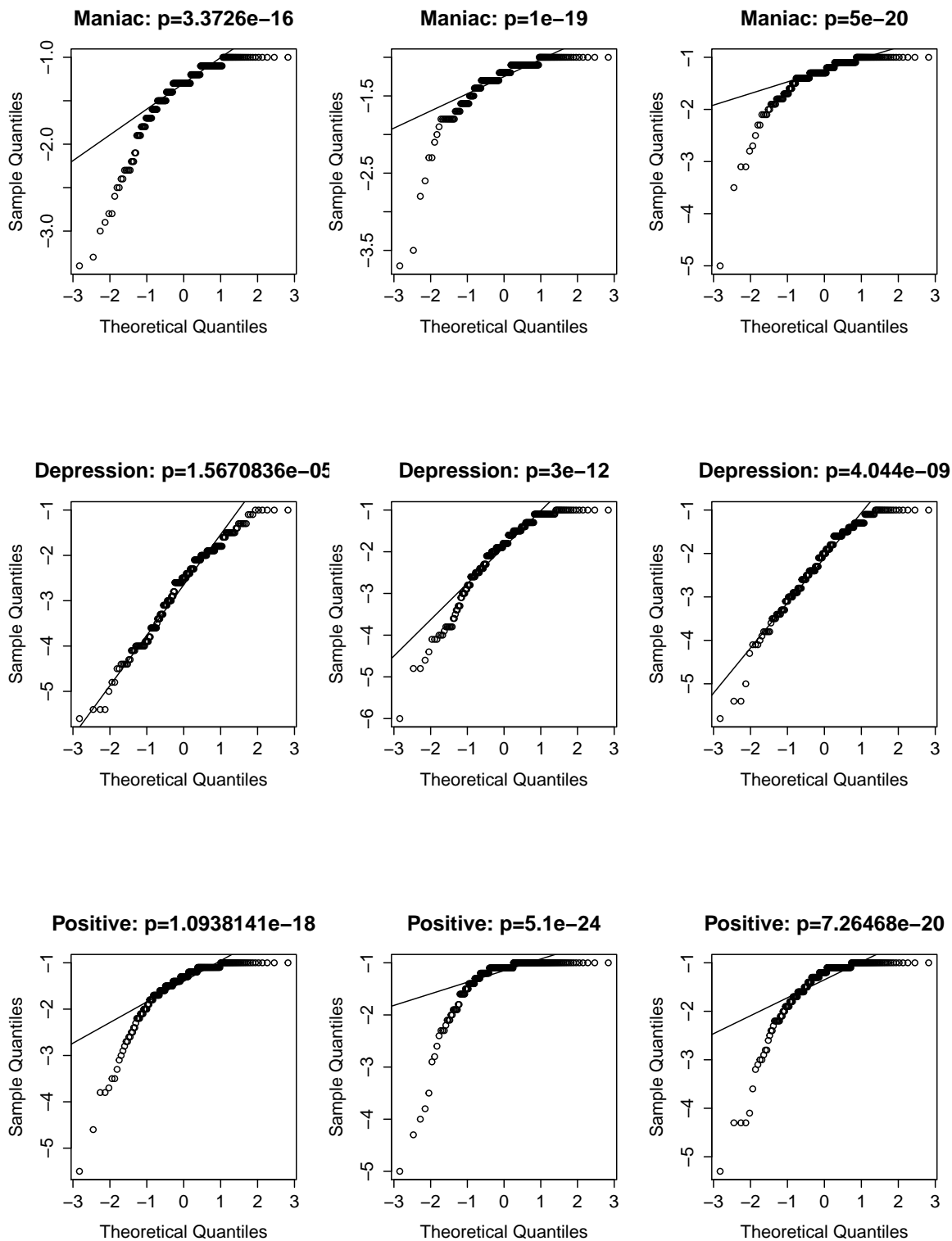
```
BPRS.Maniac<-as.numeric(BPRS.Maniac)
BPRS.Maniac.list<-split(BPRS.Maniac, group)
par(pty="s")
par(mfrow=c(1,3))
for (i in 1:3) {
  xx<-na.exclude(as.numeric(BPRS.Maniac.list[[i]]))
  p0<-shapiro.test(xx)$p.value
  qqnorm(xx, main=paste("BPRS.Maniac: p=", round(p0, 3), sep= " "))
  qqline(xx)
}
```

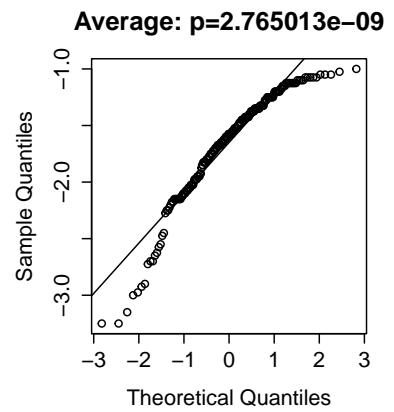
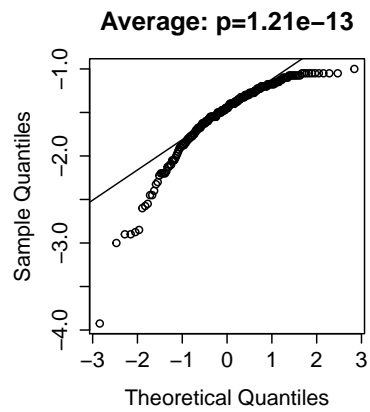
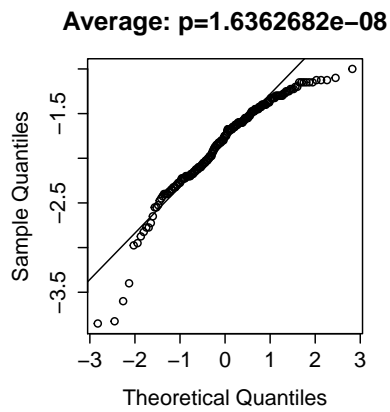
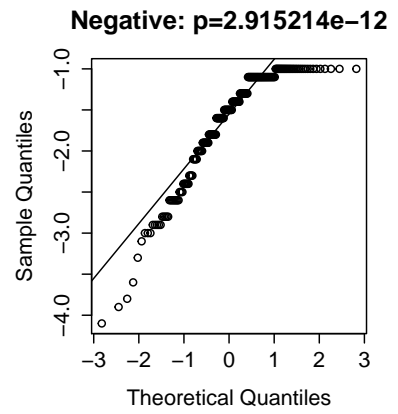
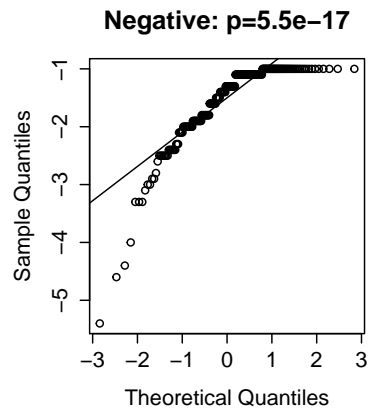
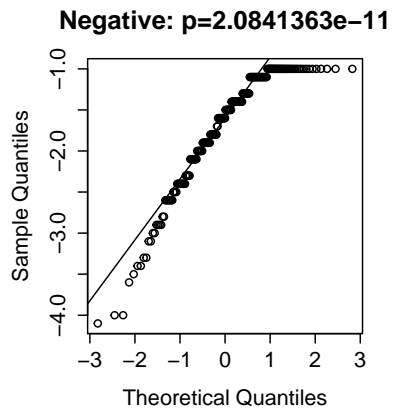
W wyniku wywołania tej funkcji wyświetlany jest wykres kwantylowy dla każdej z grup  $D^-$ ,  $D^0$  i  $D^+$  wraz z odpowiednim  $p$ -value (przedstawione na wykresie 3.3). Wynika z niego, że żadna ze zmiennych nie charakteryzuje się rozkładem normalnym, ponieważ wszystkie  $p$ -value wynoszą 0, zatem odrzucana jest hipoteza zerowa testu.

### 3.2.2. Transformacja Boxa-Coxa

Skoro rozkład zmiennych nie jest normalny, warto spróbować tak je przekształcać, żeby rozkład był jak najbardziej zbliżony do normalnego. Najczęściej stosowaną metodą jest transformacja Boxa-Coxa (zgodnie ze wzorem 1.27 z podrozdziału 1.3.1).

**Rysunek 3.3:** Wykresy kwantylowe dla kolejnych zmiennych: na osi poziomej znajdują się wartości kwantyli teoretycznych – z rozkładu normalnego, a na osi pionowej – kwantyle badanej próbki. Wykresy dla grup (od lewej):  $D^-$ ,  $D^0$  i  $D^+$





Estymacje parametru  $\lambda$  dla poszczególnych zmiennych zostały przeprowadzone w R na podstawie [23].

Poniżej zostanie przedstawione szczegółowe omówienie tej transformacji na przykładzie jednej zmiennej (BPRS.Negative), pozostałe zostały przeprowadzone w sposób analogiczny i zostaną podane same wyniki (w Tabeli 3.3).

Celem przeprowadzanych obliczeń jest znalezienie najlepiej pasującej wartości parametru  $\lambda$  takiej, że rozkład przekształconej zmiennej będzie bardziej zbliżony do rozkładu normalnego. W tym przykładzie, takiego  $\lambda$ , że rozkład zmiennej BPRS.Negative podniesionej do potęgi  $\lambda$  (lub zlogarytmowanej, gdy  $\lambda$  wynosi 0) będzie bardziej zbliżony do rozkładu normalnego.

Aby używać funkcji `boxcox` w R, należy włączyć bibliotekę `MASS`. Transformacja Boxa–Coxa wymaga, aby zmienna miała wartości dodatnie, dlatego wszystkie wartości BPRS.Negative są brane z przeciwnym znakiem (ze względu na uzyskanie monotoniczności wszystkie wartości zostały uprzednio zmienione na ujemne).

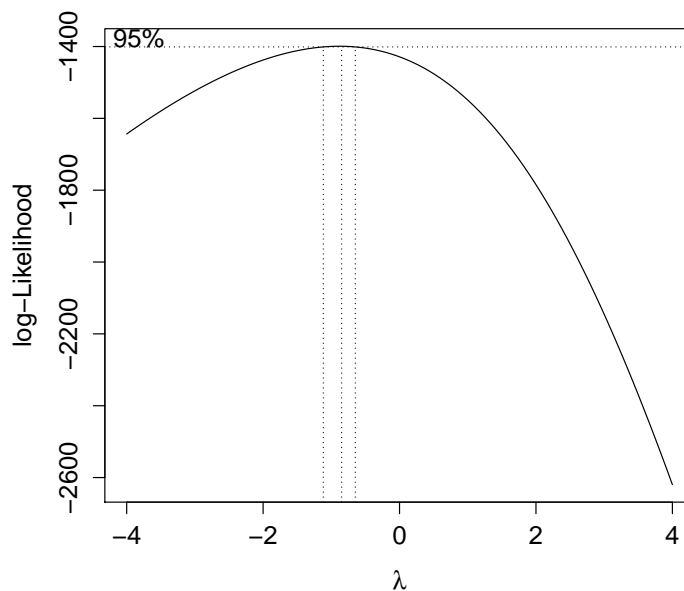
W tym przypadku wywołanie funkcji `boxcox` jest następujące

```
b=boxcox((-1)*BPRS.Negative~MANSA, data=eden, lambda=seq(-4,4,by=0.1))
```

gdzie dla parametru  $\lambda$  ustalony został przedział  $[-4,4]$  (można wybrać inny).

W wyniku ukazuje się wykres przedstawiający zależność pomiędzy wartością parametru  $\lambda$  a wartościami log–prawdopodobieństwa przy założeniu rozkładu normalnego (zależne od  $\lambda$ ), a także przedział ufności na poziomie  $\alpha=0.05$  (rys. 3.4).

**Rysunek 3.4:** Wykres zależności wartości log–prawdopodobieństwa od wartości parametru  $\lambda$  dla zmiennej BPRS.Negative.



Po wywołaniu wartości `b` wyświetlają się dwie listy: pierwsza lista to wartości  $\lambda$ , druga – odpowiadające im wartości log–prawdopodobieństwa. Żeby połączyć te listy w jedną tabelę, należy wywołać



```
lambda=b$x
lik=b$y
bc=cbind(lambda,lik)
```

Szukana wartość  $\lambda$  to argument maksymalizujący wartości log–prawdopodobieństwa, zatem po posortowaniu wartości tabeli malejąco:

```
bc[order(-lik),]
```

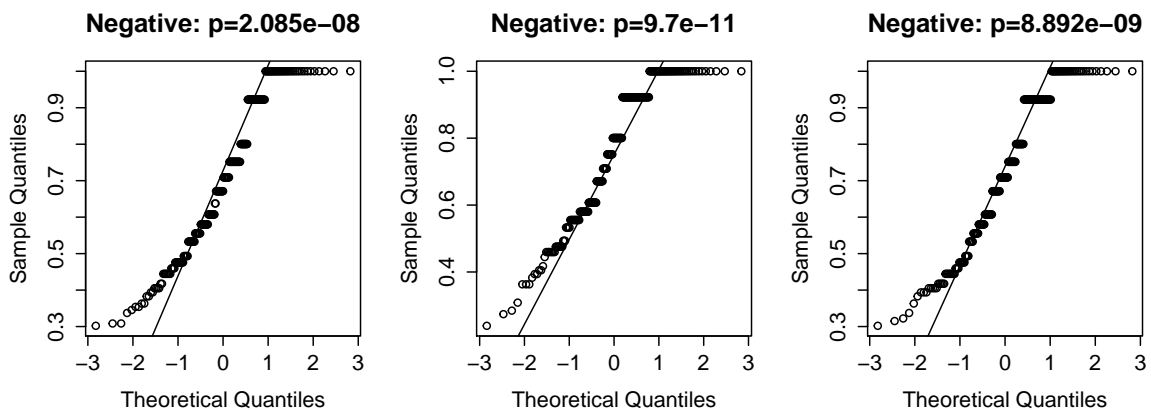
można łatwo odczytać szukaną wartość parametru  $\lambda$  (poniżej przedstawionych jest 10 pierwszych wierszy macierzy `bc` zaokrąglonych do dwóch miejsc po przecinku):

	lambda	lik
[1,]	-0.85	-1399.220
[2,]	-0.93	-1399.271
[3,]	-0.77	-1399.626
[4,]	-1.01	-1399.766
[5,]	-0.69	-1400.501
[6,]	-1.09	-1400.692
[7,]	-0.61	-1401.856
[8,]	-1.17	-1402.040
[9,]	-0.53	-1403.706
[10,]	-1.25	-1403.797

Przyjmowana jest więc wartość  $\lambda = -0.85$ .

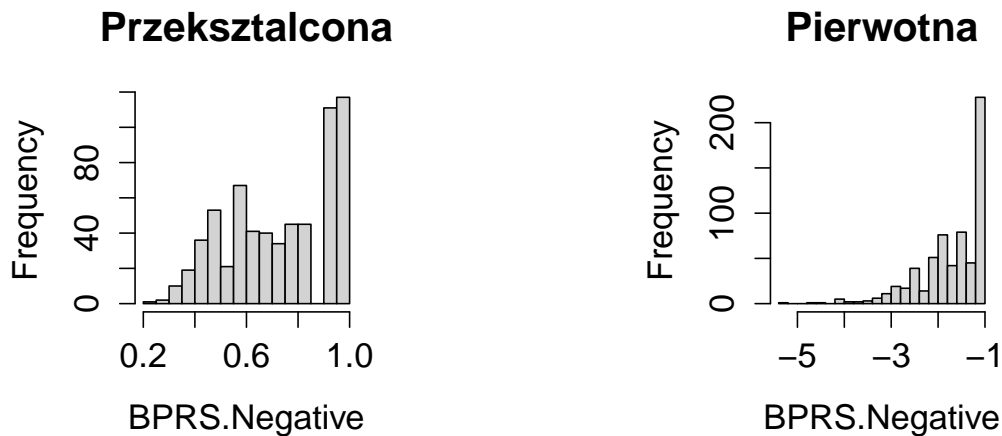
Zatem z powyższych rozważań wynika, iż zmienna `BPRS.Negative` w tej pracy powinna być podniesiona do potęgi  $-0.85$ , żeby jej rozkład był zbliżony do normalnego. Jednakże pomimo tej transformacji, test Shapiro – Wilka wykazuje brak rozkładu normalnego ( $p$ -value wynosi w przybliżeniu 0) (rys. 3.5).

**Rysunek 3.5:** Wykresy Q–Q dla trzech grup zmiennej `BPRS.Negative` (od lewej:  $D^-$ ,  $D^0$  i  $D^+$ )



Można porównać histogramy pierwotnej zmiennej `BPRS.Negative` i przekształconej transformacją Boxa–Coxa (rys. 3.6).

**Rysunek 3.6:** Histogramy zmiennej BPRS.Negative od lewej: przekształconej transformacją Boxa–Coxa, pierwotnej



Jak widać z rysunku 3.6, histogram przekształconej zmiennej wciąż odbiega od rozkładu normalnego. Lewy ogon został skrócony i pojawił się prawy ogon, jednakże można zauważyć drugi szczyt (histogram dwumodalny).

Analogiczne rozważania przeprowadzone zostały dla pozostałych zmiennych, wyniki w Tabeli 3.3.

**Tabela 3.3:** Wartości parametru  $\lambda$  transformacji Boxa–Coxa dla kolejnych zmiennych oraz wyniki testu Shapiro — Wilka dla trzech grup.

zmienna	$\lambda$	$p$ -value testu Shapiro		
		$D^-$	$D^0$	$D^+$
<i>BPRS.Maniac</i>	-2.14	0	0	0
<i>BPRS.Negative</i>	-0.84	0	0	0
<i>BPRS.Positive</i>	-2.38	0	0	0
<i>BPRS.Depression</i>	-0.07	0.003	0	0
<i>BPRS.Average</i>	-0.77	0.041	0.003	0.042

Z Tabeli 3.3 wynika, że pomimo przeprowadzenia transformacji Boxa–Coxa, żadna z przekształconych zmiennych nie ma rozkładu normalnego (zmienna BPRS.Average najbardziej się do niego zbliżyła, na co wskazuje test Shapiro–Wilka).

Część testów dla zbiorów klasyfikatorów przeprowadza się przy założeniu, że dane w poszczególnych grupach mają rozkład normalny, zbliżony do niego lub można ich rozkład poprzez transformację monotoniczną sprowadzić do rozkładu normalnego. W podrozdziale 3.2 w bardzo dokładny sposób przedstawione zostało to, że dla wykorzystywanych danych założenie

to nie jest spełnione. Założenie o wymaganej dystrybucji rozkładu wykorzystywane jest w przypadku posługiwania się statystyką testową, badającą różnice wariancji dla określonej liczby grup. Dotyczy to przede wszystkim testu t-Studenta czy Z-testu. W przypadku danych medycznych, gdy rozmiar próby jest bardzo duży, a zmienne są niezależne, dystrybucja rozkładu okazuje się nie być już tak ważna dzięki temu, że zachodzi centralne twierdzenie graniczne. Dowód powyższego znaleźć można w [14]. W artykule [14] podaje się, że próbę uznaje się za dużą dla liczby obserwacji większej od 100. W przypadku danych o stanie psychicznym pacjentów w każdej z grup  $D^-$ ,  $D^0$  i  $D^+$  wymóg ten jest spełniony. W pracy tej dla otrzymania jak najdokładniejszych wyników, jak najlepiej dopasowanych do danych wszędzie tam, gdzie jest to możliwe, estymacja przeprowadzona zostanie przy pomocy metod nieparametrycznych, natomiast w pozostałych wykorzystane zostaną testy parametryczne, których wyniki również uznać można za poprawne.

### 3.3. Porównanie własności pojedynczych markerów

W powyższych rozważaniach stwierdzono, że rozkład zmiennych nie jest rozkładem normalnym. W tej części pracy wykorzystana zostaje zatem estymacja nieparametryczna opisana w Rozdziale 1, która umożliwia przybliżanie rozkładu normalnego. Użyte będą pierwotne dane (nie przekształcone transformacją Boxa–Coxa).

#### Estymacja parametrów VUS dla klasyfikatorów na podstawie danych o stanie psychicznym pacjentów Europy

W wykorzystywanych danych każdemu pacjentowi przypisana została wartość sześciu wskaźników odzwierciedlających stan jego zdrowia psychicznego. W części tej zostanie przedstawiona próba analizy i oceny przydatności i poprawności pięciu z tych klasyfikatorów do przyporządkowywania pacjentów do odpowiednich dla nich grup, które zostały z góry określone przy wykorzystaniu szóstego klasyfikatora. Celem jest bowiem znalezienie miary, czy cechy, której analiza mogłaby posłużyć do stworzenia testu diagnostycznego.

Jak zostało to już przedstawione w części teoretycznej, każdy test diagnostyczny charakteryzuje VUS, czyli łączna miara poprawności diagnostycznej, w przypadku, gdy rozróżnia się trzy stany.

W poprzednim rozdziale zostały opisane podstawowe statystyki na temat zmiennych, a przeprowadzony test Shapiro–Wilka potwierdził, że dane dla żadnego markera i dla żadnej z grup nie mają rozkładu normalnego. Dostępne funkcje pakietu `DiagTest3Grp` umożliwiają jednak przeprowadzenie estymacji nieparametrycznej, która to właśnie została wykorzystana. W przypadku miary VUS estymacja nieparametryczna wykorzystuje metodę bootstrap.

Przedstawiony poniżej kod ukazuje użycie funkcji VUS dla zmiennej `BPRS.Maniac`:

```
BPRS.Negative<-as.numeric(eden$BPRS.Negative)
BPRS.Negative.list<-split(BPRS.Negative, group)
xx <- BPRS.Negative.list$"D-"
yy <- BPRS.Negative.list$"D0"
zz <- BPRS.Negative.list$"D+"
vus.nonpar<-VUS(x=xx, y=yy, z=zz,method="NonPar", NBOOT=50)
vus.nonpar
par(pty="m")
plot(vus.nonpar)
```

W ten sam sposób, zmieniając jedynie nazwę postępuje się dla pozostałych klasyfikatorów. Zwrócone zostają wartości poszczególnych charakterystyk tej zmiennej:

The DiagTest3Grp summary measure: VUS		
Method used for VUS:NonPar		
Raw Data Summary:		
	n	mu sd
D-	211	-1.7236 0.68737
D0	209	-1.6727 0.66706
D+	222	-1.5725 0.69112
VUS=0.2461, 95% CI=0.2289~0.2665		
Best cut-points: lower=-1.9, upper=-1.1		
The group correct classification probabilities are:		
	Sp	Sm Se
	0.3697	0.3349 0.4279

Schemat jest zachowany dla każdej ze zmiennych. W celu przejrzystości tekstu oraz łatwiejszego porównania pomiędzy badanymi klasyfikatorami, wszystkie wyniki zostały przedstawione w Tabeli 3.4. Wszystkie oznaczenia tabeli zwracanych wyników wprowadzone w Rozdziale 1 zostają zachowane.

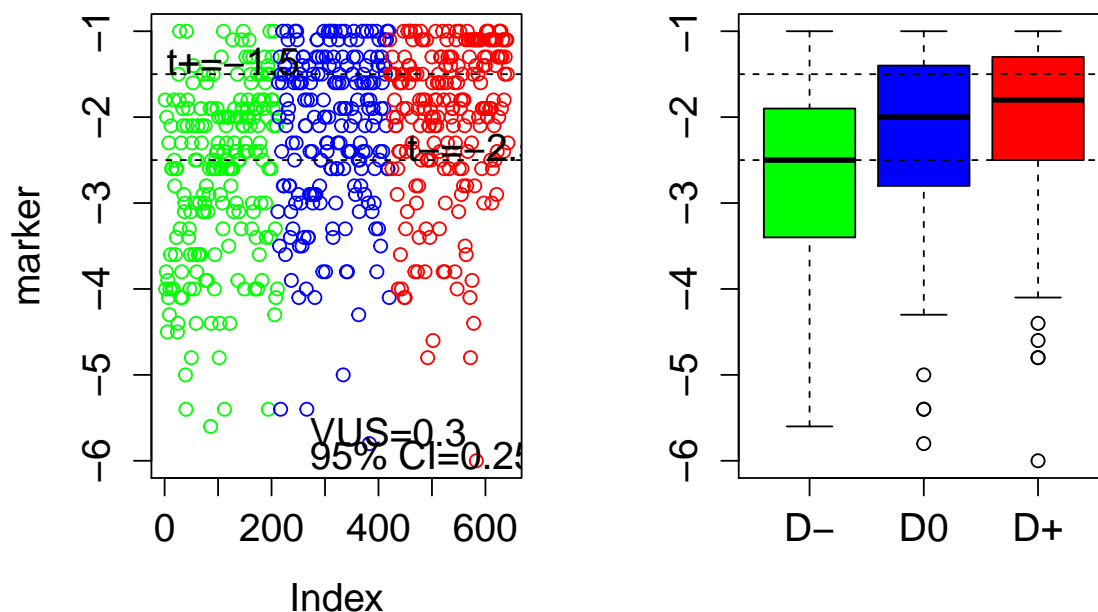
Jednocześnie dla każdego klasyfikatora zwracany zostaje wykres pudełkowy oraz wykres rozrzutu zmiennej dla podziału na trzy grupy, wraz z zaznaczonymi wartościami punktów  $t_-$  i  $t_+$ .

**Tabela 3.4:** Wyniki estymacji miary VUS

	<i>Maniac</i>	<i>Negative</i>	<i>Positive</i>	<i>Average</i>	<i>Depression</i>
VUS	25.38%	24.61%	30.44%	28.83%	29.74%
95 CI	22.52–28.11%	22.89–26.65%	27.28–32.75%	25.02–31.71%	26.07–32.85%
$t_-$	-1.5	-1.9	-1.4	-1.85	-2.5
$t_+$	-1.2	-1.1	-1	-1.4	-1.5
<i>Sp</i>	.3175	.3697	.4645	.4408	.5261
<i>Sm</i>	.3014	.3349	.4019	.3876	.3445
<i>Se</i>	.5541	.4279	.4019	.4775	.4009

Na podstawie obliczeń widać, że każdy z markerów charakteryzuje się poprawnością w różnieniu grup lepszą od przypadkowego wyboru (dla każdej zmiennej  $VUS > \frac{1}{6}$ ). Żadnego więc z markerów nie można uznać za kompletnie bezużyteczny. Niestety, wartości funkcji VUS są na tyle małe, że w przypadku realnie wykorzystywanych testów diagnostycznych mogłoby się okazać, że zwracają one zbyt wiele wyników niepoprawnych. Zauważalne jest także pewne podobieństwo zarówno dla samych wartości VUS, ale i dla pozostałych wartości. Punkty, które najlepiej rozdzielają pacjentów chorych, zdrowych i tych w początkowym stadium choroby

Rysunek 3.7: Wykres rozproszenia oraz wykres pudełkowy BPRS.Depression

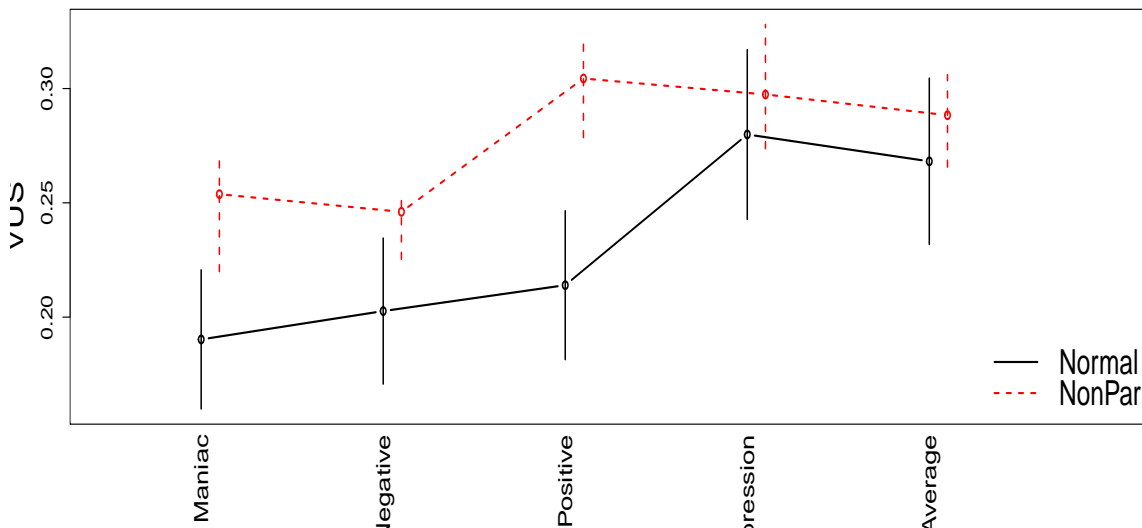


przyjmują również podobne wartości. Przy zastosowaniu tych punktów oddzielających i wykorzystaniu któregoś z markerów jako testu diagnostycznego widać, że marker BPRS.Positive ma najlepszą zdolność do poprawnego klasyfikowania osób pochodzących z dowolnej grupy — prawdopodobieństwo poprawnego zdiagnozowania dowolnej osoby wynosi powyżej 40%. W przypadku zmiennych BPRS.Maniac i BPRS.Depression wyższe prawdopodobieństwo odpowiedniego rozpoznania pacjentów chorych lub zdrowych jest związane z większą szansą na błąd podczas badań pozostałych osób. Okazuje się więc, że wskaźnik BPRS.Positive ma najlepszą spośród dostępnych zdolność diagnostyczną. Aczkolwiek przy konstrukcji testu może rodzić się również pytanie, czy nie warto cech tych stosować niejako wymiennie, tzn. czy dla pewnej grupy testów nie jest ważniejszym poprawne rozpoznawanie wszystkich pacjentów w zaawansowanym stadium choroby, kosztem popełnianie systematycznie większego błędu w diagnozowaniu pozostałych grup pacjentów, o ile oczywiście jego rząd wielkości będzie się mieścił w określonym przedziale.

Wszystkie wyniki wskaźnika VUS dla estymacji parametrycznej oraz nieparametrycznej przedstawione zostały także na rysunku 3.8. Linia przerywaną, czerwoną, zaznaczono wartości tego wskaźnika dla metod nieparametrycznych, natomiast ciągłą, czarną, dla parametrycznych. Pionowe odcinki dla każdego zaznaczonego punktu odpowiadają zadanym przedziałom ufności wartości VUS.

Jak widać na rysunku 3.8, różnice pomiędzy wartościami osiąganymi obiema metodami nie są znaczne, największa w przypadku zmiennej BPRS.Positive — wynosi około 10 punktów procentowych.

**Rysunek 3.8:** Wartość wyestymowanej miary VUS przy pomocy metody parametrycznej i nieparametrycznej dla wszystkich markerów.



### Indeks Youdena każdej ze zmiennych

Rozszerzony indeks Youdena został opisany w podrozdziale 1.3. Przybliżone zostały w nim metody zarówno parametryczne, jak i nieparametryczne. Skoro w tym przypadku założenie normalności rozkładu nie jest spełnione, użyta została estymacja nieparametryczna.

Wszystkie obliczenia zostały wykonane korzystając z funkcji `Youden3Grp` z biblioteki `DiagTest3Grp`.

Jak było opisywane wcześniej, funkcja `Youden3Grp` może przyjmować jedną z czterech z określonych metod estymacji nieparametrycznej, a mianowicie jedną z poniższych metod:

**TN** — metoda ta implementuje transformację Boxa–Coxa do zmiennej (czyli wylicza najlepszy parametr  $\lambda$  i następnie przekształca zmienną zgodnie ze wzorem 1.27), po czym traktuje zmienną tak, jakby miała rozkład normalny;

**EMP** — metoda polega na empirycznym przybliżeniu dystrybuanty i wyliczeniu indeksu zgodnie ze wzorem 1.29;

**KS** — w tej metodzie używane jest jądrowe wygładzanie Gaussa dystrybuanty rozkładu zmiennej przy ustalonej szerokości okna wynoszącej

$$h_i = 1.06n_i^{-0.2} \min \left\{ \hat{\sigma}_i, \frac{IQR_i}{1.34} \right\},$$

$i = -, 0, +$ , gdzie  $IQR_i$  oznacza rozstęp ćwiartkowy w każdej z trzech grup (czyli różnicę pomiędzy trzecim a pierwszym kwantylem), a  $\hat{\sigma}_i$  oznacza odchylenie standardowe;

**KS-SJ** — metoda ta jest zbliżona do metody KS, ale aplikowany jest tutaj sposób wyliczania szerokości okna metodą autorstwa Sheathera i Jonesa (opisaną np. w [18], [19]). Szerokość ta w pakiecie `DiagTest3Grp` obliczana jest za pomocą funkcji z pakietu `KernSmooth`.

Użycie tej funkcji zostanie zaprezentowane na przykładzie jednej zmiennej (BPRS.Negative), wyniki dla pozostałych zmiennych otrzymywane są w analogiczny sposób.

Wywołanie funkcji `Youden3Grp` musi zostać poprzedzone zdefiniowaniem oddzielnych list wartości zmiennej `BPRS.Negative` w rozróżnieniu na poszczególne podgrupy (czyli następuje rozdzielanie wartości na trzy grupy, nazwane `xx`, `yy`, `zz` dla odpowiednio  $D^-$ ,  $D^0$  i  $D^+$ ). W R kod generujący te trzy listy zaprezentowany jest poniżej.

```
BPRS.Negative<-as.numeric(eden$BPRS.Negative)
BPRS.Negative.list<-split(BPRS.Negative, group)
xx <- BPRS.Negative.list$"D-"
yy <- BPRS.Negative.list$"D0"
zz <- BPRS.Negative.list$"D+"

```

Po rozgrupowaniu wartości, można przystąpić do wywołania właściwej funkcji `Youden3Grp`:

```
youden<-Youden3Grp(x=xx, y=yy, z=zz, method="EMP", lam.plus=222/642,
+ lam0=209/642, lam.minus=211/642)
youden

```

gdzie polecenie `youden` powoduje wyświetlenie na ekranie wyników. W tej funkcji została użyta metoda `EMP`, czyli empirycznego przybliżania dystrybuant dla poszczególnych grup.

Zwrócone zostają wartości poszczególnych charakterystyk tej zmiennej:

```
The DiagTest3Grp summary measure:  Youden

Method used for Youden:EMP

Raw Data Summary:
      n      mu      sd
D- 211 -1.723697 0.6873743
D0 209 -1.672727 0.6670635
D+ 222 -1.572523 0.6911280

Youden=0.06,95% CI=-0.0037~0.0903

Best cut-points: lower=-1.9346, upper=-1.4689

The group correct classification probabilities are:
      Sp      Sm      Se
0.3081  0.2488  0.5631

```

Można odczytać, że wartość rozszerzonego indeksu Youdena dla zmiennej BPRS.Negative przy użyciu tej metody wynosi 6%. Wnioskować można więc, że ta zmienna nie stanowi dobrego wskaźnika rozróżniającego pomiędzy grupami (wynik ten jest blisko zera, które oznacza przyporządkowanie nie lepsze niż przez przypadek).

Wyświetlane są też optymalne punkty przecięcia: dolny próg —  $t_-$ , wynosi -1.93, a górny —  $t_+$ , wynosi -1.47, a także wartości prawdopodobieństw odpowiedniego przypisania do danej grupy ( $Sp$ ,  $Sm$ ,  $Se$  kolejno dla  $D^-$ ,  $D^0$  i  $D^+$ ). Można sprawdzić, że wartość rozszerzonego indeksu Youdena rzeczywiście wyrażona jest poprzez wartości tych prawdopodobieństw, a mianowicie

$$J(t_-, t_+) = \frac{1}{2} (Sp(t_-) + Sm(t_-, t_+) + Se(t_+) - 1),$$

czyli w tym przypadku

$$\begin{aligned} J(-1.9346, -1.4689) &= \frac{1}{2} (Sp(-1.9346) + Sm(-1.9346, -1.4689) + Se(-1.4689) - 1) = \\ &= (0.3081 + 0.2488 + 0.5631 - 1) = 0.06, \end{aligned}$$

co jest zgodne z wyświetlonym wynikiem przy użyciu funkcji `Youden3Grp`.

Implementacja pozostałych metod przebiega analogicznie do powyższego przykładu, szczególne wyniki dla BPRS.Negative i pozostałych markerów zamieszczone są w Tabeli 3.5.

Jak wynika z wartości znajdujących się w Tabeli 3.5, rozszerzony indeks Youdena ma dosyć niskie wartości (od 5.27% do 16.08%). Nie ma jednej dominującej metody, która dawałaby najlepsze rezultaty. Co więcej, niektóre prawdopodobieństwa  $Sm$ , czyli prawdopodobieństwa, że mając wylosowany jeden z wyników próbki z grupy  $D^0$  poprawnie zostanie on zaklasyfikowany do tejże grupy, przyjmują wartości ujemne. Jest to powodowane tym, że w tych estymacjach punkt  $t_+$  jest mniejszy niż  $t_-$ , wbrew wcześniejszym założeniom. Przypuszczalnie jest to błąd w funkcji `Youden3Grp`, ponieważ takie wyniki nie powinny być dopuszczalne. Poruszona w podrozdziale 1.3.1 możliwość braku spełnienia nierówności  $t_-^* < t_+^*$  dla optymalnych punktów progowych może być przyczyną tego typu problemów. Najprawdopodobniej funkcja w omawianym pakiecie nie sprawdza tej zależności. Problem ten został zgłoszony autorom pakietu.

Największe wartości (a zatem najlepsze możliwości rozróżniania) mają zmienne BPRS.Depression (indeks Youdena na poziomie 15%) oraz BPRS.Average (na poziomie 14.5%). Maksymalna wartość rozszerzonego indeksu Youdena wynosi 16.08% (dla zmiennej BPRS.Depression). Są to jednakże niskie wyniki — wskazują, że najprawdopodobniej żadna ze zmiennych nie ma własności dobrego rozróżniania pomiędzy grupami.

Wartości z Tabeli 3.5 można przedstawić za pomocą wykresu (rys. 3.9). Można go wywołać dzięki jednej z funkcji zamieszczonych w pakiecie `DiagTest3Grp`.

W tym celu należy najpierw wywołać skrypt `AD.analysis` (zamieszczony jako załącznik do pakietu), a następnie polecenie

```
all.youden <- AD.analysis(edn, type = "Youden", NBOOT = 100,
+ randomStart.N=10)
```

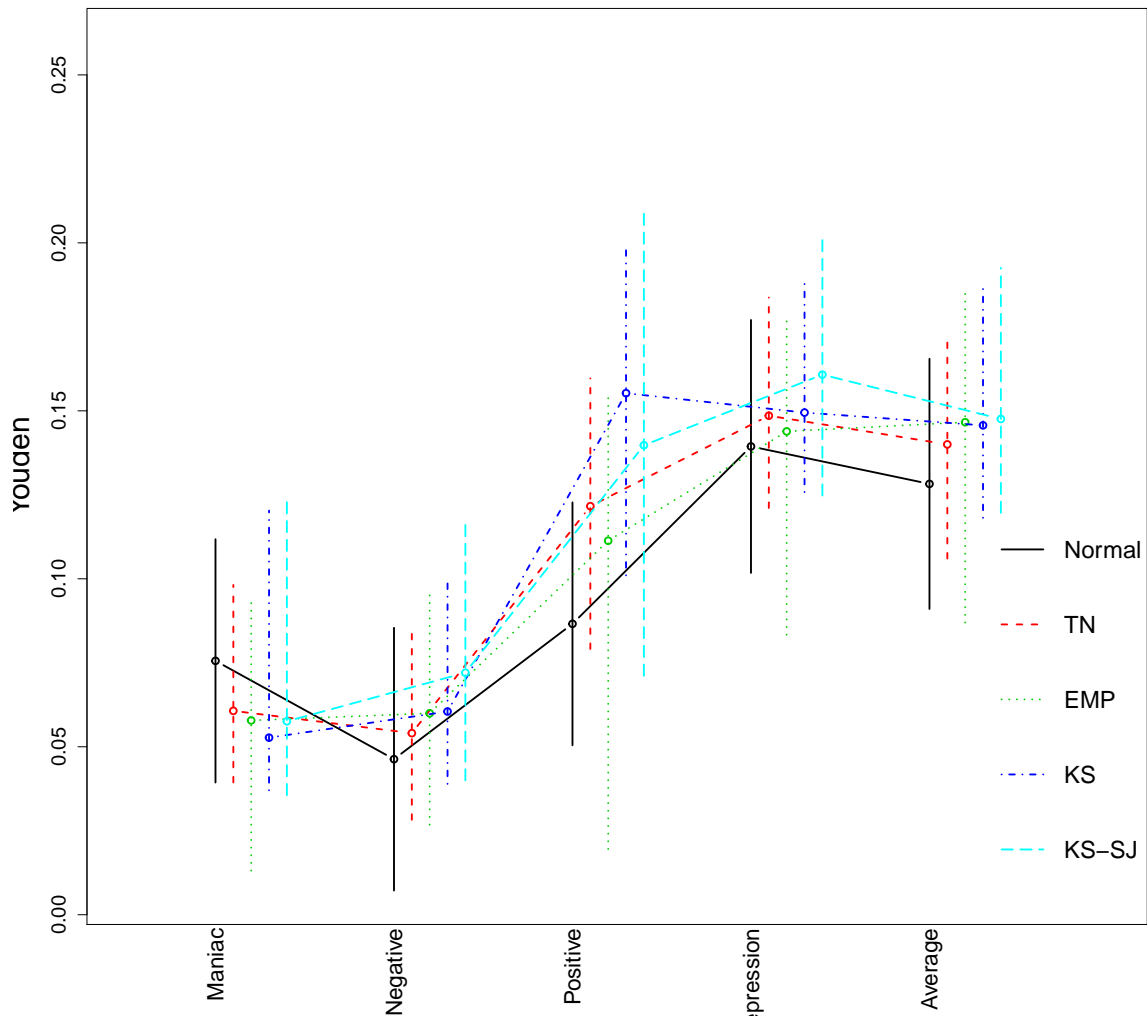


**Tabela 3.5:** Wartości rozszerzonego indeksu Youdena, optymalne punkty przecięcia  $t_-$  (dolny) i  $t_+$  (górny) oraz wartości prawdopodobieństwa odpowiedniego przypisania do poszczególnych grup w podziale na różne metody estymacji nieparametrycznej dla każdej ze zmiennych.

BPRS.Negative						
metoda	Youden	$t_-$	$t_+$	sp	sm	se
TN	5.4%	-1.5912	-1.543	49.05%	36.4%	58.12%
EMP	6%	-1.9346	-1.4689	30.81%	24.88%	56.31%
KS	6.05%	-1.561	-1.1349	51.32%	13.01%	47.77%
KS-SJ	7.2%	-1.5729	-1.2334	49.99%	21.55%	42.87%
BPRS.Positive						
TN	12.16%	-1.2472	-1.3066	61.90%	-7.46%	69.88%
EMP	11.13%	-1.312	-1.6764	46.45%	-12.92%	88.74%
KS	15.53%	-1.2439	-1.1965	58.44%	5.5%	46.11%
KS-SJ	13.98%	-1.1798	-1.4285	63.6%	-18.23%	82.57%
BPRS.Maniac						
TN	6.07%	-1.312	-1.3907	48.32%	-10.26%	74.07%
EMP	5.78%	-1.1948	-1.6425	67.3%	-44.02%	88.29%
KS	5.27%	-1.1717	-1.8042	66.39%	-49.18%	93.33%
KS-SJ	5.76%	-1.1447	-1.6778	69.85%	-47.27%	88.93%
BPRS.Depression						
TN	14.85%	-2.2036	-2.0295	64.22%	8.29%	57.19%
EMP	14.39%	-2.5401	-2.2401	47.87%	12.44%	68.47%
KS	14.95%	-1.9142	-2.2424	73.98%	-11.76%	67.68%
KS-SJ	16.08%	-1.8312	-2.2139	77.89%	-13.41%	67.67%
BPRS.Average						
TN	14%	-1.637	-1.5843	63.02%	3.03%	61.94%
EMP	14.66%	-1.8334	-1.7069	44.08%	9.57%	75.68%
KS	14.57%	-1.725	-1.6049	51.43%	11.87%	65.84%
KS-SJ	14.76%	-1.7676	-1.5934	49.73%	14.3%	65.48%

które powoduje, że wyświetla się wykres dla każdego markera i każdej metody przedstawiający wartości rozszerzonego indeksu Youdena i przedziały ufności (pionowe odcinki).

**Rysunek 3.9:** Analiza wszystkich markerów w kontekście rozszerzonego indeksu Youdena



Podobnie jak w przypadku miary VUS, wyniki estymacji rozszerzonego indeksu Youdena przy pomocy metod parametrycznych i nieparametrycznych są do siebie zbliżone. Warto również zauważyć, że i w tym przypadku zmienna BPRS.Positive charakteryzuje się największą rozbieżnością pomiędzy wartościami osiąganymi tymi pięcioma metodami.

### 3.4. Analiza wszystkich markerów — porównanie

Mając do dyspozycji więcej czynników, których możliwość zastosowania w testach diagnostycznych dla trzech grup jest analizowana, warto jest znaleźć spośród nich te, które „niosą” tę samą informację oraz te, które być może mogą się wzajemnie uzupełniać. W tym celu można porównywać ich miary VUS i Youdena. Warunkiem jest, aby w przypadku porównywania markerów estymacja powyższych wartości odbywała się na tej samej próbie.

Pierwszą analizą w przypadku obserwacji wielu markerów może być przeprowadzenie testu zbiorowego. Jego kod wywołujący oraz wyniki zostały umieszczone w poniższych ramkach. Pierwsza z nich dotyczy miary VUS, druga rozszerzonego Indeksu Youdena.

```
new.eden <- data.frame(group=group, eden)
omni.vus <- DiagTest3Grp.Test(dat=new.eden, paired=TRUE,
type = "VUS", p=0, q=0, mu=0, conf.level=0.95, alternative="two.sided")
omni.vus
```

```
Chi-Square test
data: Test new.eden on VUS
Chi-square = 4658.798, df = 5, p-value < 2.2e-16
alternative hypothesis: true diff in VUS is not equal to 0
percent confidence interval:
NA NA
sample estimates:
VUS of BPRS.Maniac      VUS of BPRS.Negative
      .1902              .2026
VUS of BPRS.Positive    VUS of BPRS.Depression    VUS of BPRS.Average
      .2139              .2799              .2681
```

```
new.eden <- data.frame(group=group, eden)
omni.youden <- DiagTest3Grp.Test(dat=new.eden, paired=TRUE,
type = "Youden", p=0, q=0, mu=0, conf.level=0.95, alternative="two.sided")
omni.youden
```

```
Chi-Square test
data: Test new.eden on Youden
Chi-square = 2555.88, df = 5, p-value < 2.2e-16
alternative hypothesis: true diff in Youden is not equal to 0
percent confidence interval:
NA NA
sample estimates:
Youden of BPRS.Maniac    Youden of BPRS.Negative
      .0755              .0463
Youden of BPRS.Depression Youden of BPRS.Average    Youden of BPRS.Positive
      .1393              .1282              .0865
```

**Tabela 3.6:** Statystyka testowa  $Z$  oraz  $p$ -value dla porównywania parami objętości pod powierzchnią ROC

Marker	Row	Negative	Positive	Depression	Average
Maniac	Z-statistic	-.5580	-1.2600	-3.9680	-4.1230
	raw P	.5770	.2080	7.24e-05	3.73e-05
	adjusted P	.5770	.26	.00014	8e-05
Negative	Z-statistic		-.6080	-3.7350	-4.5050
	raw P		.543	.00019	6.65e-06
	adjusted P		.5770	.00028	1.66e-05
Positive	Z-statistic			-3.0840	-3.8530
	raw P			.0021	.00012
	adjusted P			.00278	.00019
Depression	Z-statistic				.8880
	raw P				.3740
	adjusted P				.4320

W obu przypadkach wartość  $p$ -value statystyki testowej jest mniejsza od zadanego poziomu istotności  $\alpha = 0.05$ , wobec tego zarówno miara VUS jak i indeks Youdena nie są ekwiwalentne wśród wszystkich rozważanych markerów, dlatego mają one odmienne poziomy zdolności rozróżniania pacjentów zmagających się ze schorzeniami psychicznymi.

Następna analiza polega na porównywaniu wszystkich możliwych par. Markerów jest pięć, zatem należy wykonać dziesięć porównań miar Youden lub VUS. Do tego celu wykorzystana zostanie funkcja dokładniej opisana w podrozdziale 2.2. Jej implementacja w przypadku miary VUS to:

```
pairwise.vus <- Pairwise.DiagTest3Grp.Test(new.eden, paired = TRUE,
type = "VUS", p=0, q=0, conf.level=0.95, p.adjust.method = "fdr")
print(pairwise.vus$print.matrix[1:15, 1:7], na = "")
```

W celu porównań rozszerzonego indeksu Youdena argument typu został oznaczony jako **Youden**.

Wyniki obu porównań zostały umieszczone w Tabelach 3.6 i 3.7.

Te same wyniki przedstawiane są jednocześnie w postaci mapy ciepła, której wartościami są  $p$ -value dopasowane. Dla analizowanych danych o stanie psychicznym pacjentów są to Rys. 3.10 oraz Rys. 3.11.

W przypadku miary VUS statystycznie rozróżnialne nie są BPRS.Maniac i BPRS.Negative oraz BPRS.Negative i BPRS.Positive. Jednocześnie dla tych par to odpowiednio BPRS.Maniac oraz BPRS.Positive odznaczały się wyższą „zdolnością” poprawnego przyporządkowania każdej możliwej trójki obserwacji. Powyższe porównanie parami wskazuje również, że i te dwa markery nie różnią się od siebie statystycznie, zatem ponownie spośród tych dwóch można wybrać ten lepszy, tzn. „popołniający” mniej błędów, czyli BPRS.Positive.

Dla indeksu Youdena statystycznie nierozróżnialne są BPRS.Maniac i BPRS.Positive oraz BPRS.Depression i BPRS.Average. Wyniki dotyczące tej miary zawarte w Tabeli 3.5 wskazują, że wśród tych par trudnym jest wybranie tego najlepszego, ze względu na błędy, które pojawiły się podczas implementacji testów — ujemne prawdopodobieństwo.

Przy okazji porównywania markerów ze sobą w ramach jednego wskaźnika warto zobaczyć, czy dla tego samego klasyfikatora statystyki liczone wraz z miarą VUS oraz indeksem Youdena

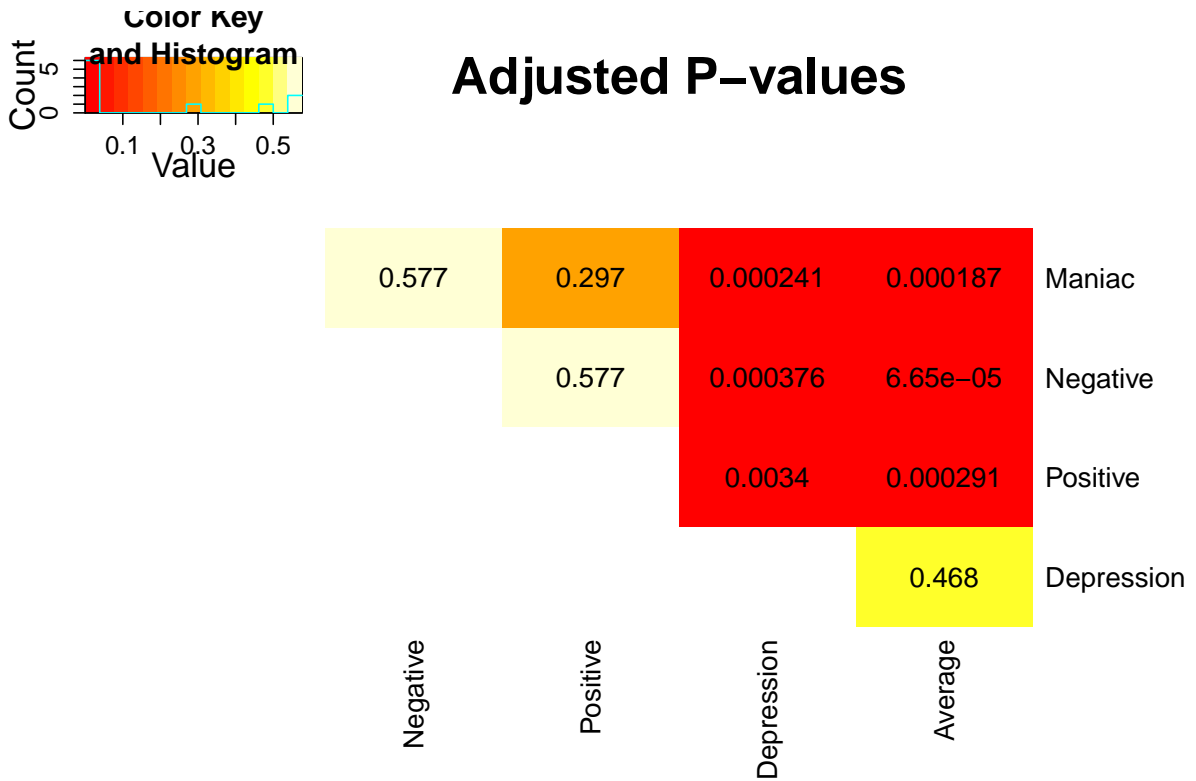
**Tabela 3.7:** Statystyka testowa  $Z$  oraz  $p$ -value dla porównywania parami rozszerzonego indeksu Youdena

Marker	Row	Negative	Positive	Depression	Average
Maniac	Z-statistic	1.0860	-.4990	-2.5900	-2.4290
	raw P	.2780	.6180	.0096	.0152
	adjusted P	.3200	.6180	.0160	.0227
Negative	Z-statistic		-1.7790	-4.0770	-4.9390
	raw P		.0752	4.57e-05	7.85e-07
	adjusted P		.0940	9.79e-05	1.96e-06
Positive	Z-statistic			-2.3640	-2.6180
	raw P			.0181	.0088
	adjusted P			.0247	.0160
Depression	Z-statistic				.8870
	raw P				.3750
	adjusted P				.4020

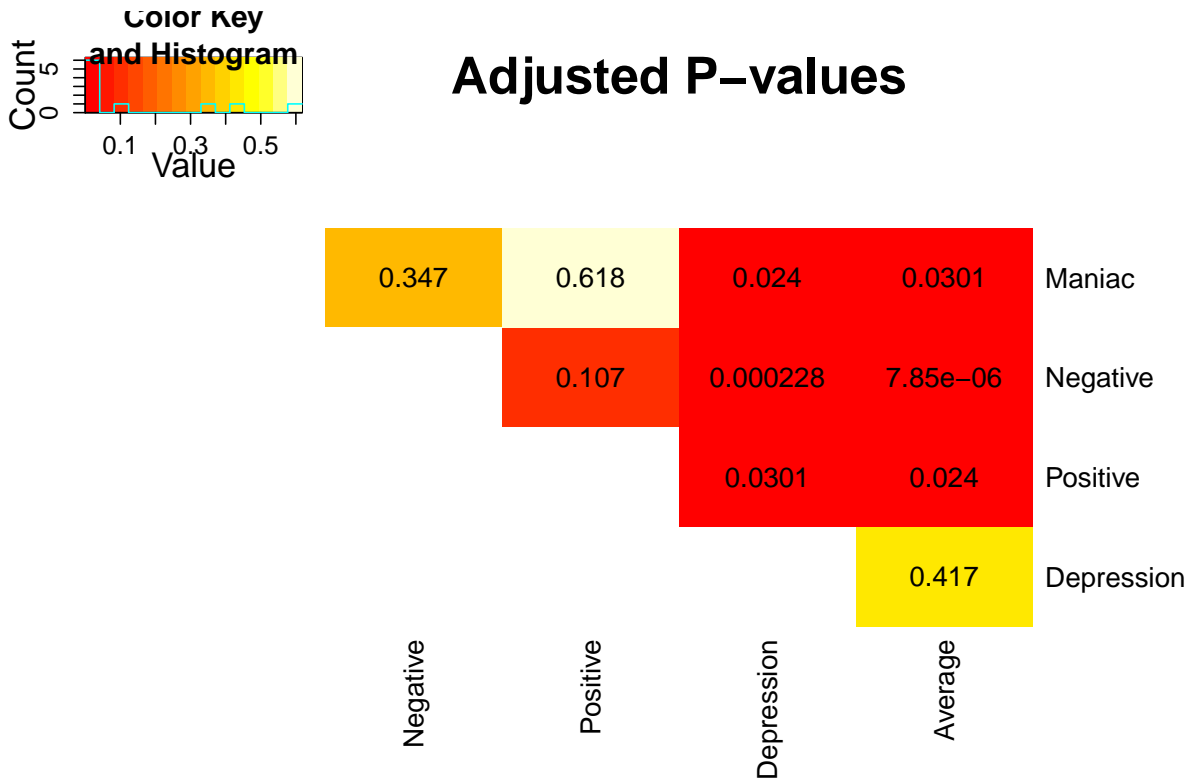
dostarczają podobnych informacji o możliwości zastosowania ich w testach diagnostycznych. Oczywiście porównaniu nie będą poddane te dwie miary, ale wartości  $Sp$ ,  $Sm$ ,  $Se$ .

Dla każdej ze zmiennych wartości te znacznie się różnią. Szczególnie rozpatrując  $Sm$ , czyli prawdopodobieństwo prawidłowego przypisania pacjentów we wczesnym stadium choroby do grupy  $D^0$ , różnice te są największe. W przypadku wyników dla rozszerzonego indeksu Youdena,  $Sp$  oraz  $Se$  mają stosunkowo duże wartości, przy niskich prawdopodobieństwach  $Sm$ . Dla miary VUS przyjmują one zbliżone do siebie wartości.

**Rysunek 3.10:** Mapa ciepła  $p$ -value dopasowanego dla sparowanych klasyfikatorów: VUS.



**Rysunek 3.11:** Mapa ciepła  $p$ -value dopasowanego dla sparowanych klasyfikatorów: indeks Youdena.



## Rozdział 4

# Podsumowanie

Praca ta podejmuje tematykę wykorzystywanych w medycynie testów diagnostycznych. Szczególna uwaga zostaje poświęcona testom dla trzech grup. Poruszone zostały aspekty nie tylko teoretyczne, ale również praktyczne zastosowanie dostępnych technik i pakietu **DiagTest3Grp** do analizy własności odpowiedniego przyporządkowania do trzech grup dostępnych markerów.

Głównym celem niniejszej pracy była analiza rzeczywistych danych medycznych oraz próba wskazania takiej zmiennej dotyczącej zachowań psychopatologicznych, którą uznać można za najlepiej różnicującą jakość życia pacjentów cierpiących na choroby psychiczne.

Opracowywane dane dotyczące chorób psychicznych zostały podzielone na trzy grupy względem zmiennej wyrażającej jakość życia (niska, średnia i wysoka jakość życia). Następnie za pomocą funkcji pakietu **DiagTest3Grp** wyliczone zostały wskaźniki poprawności testu diagnostycznego, których wyniki mogą świadczyć o tym, że kierując się wartościami danego markera można w jak najlepszy sposób przyporządkować obserwacje do zdefiniowanych wcześniej trzech grup. Głównymi miarami określającymi zdolności zmiennych do takiej klasyfikacji są VUS oraz rozszerzony indeks Youdena. Stanowią one prawdopodobieństwa tego, że obserwacje zostaną odpowiednio przypisane do danej grupy. Im wyższe ich wartości, tym lepsze zdolności markera do rozróżniania.

W przypadku estymacji obu miar wykorzystywane są statystyki, które wymagają spełnienia warunku o normalności rozkładu danego markera w każdej z grup. W tym przypadku dane nie charakteryzowały się takim rozkładem, dlatego też w celu otrzymania dokładniejszych wyników, wszędzie tam, gdzie to było możliwe, przeprowadzona została estymacja nieparametryczna. Ze względu jednak na bardzo dużą liczbę obserwacji, rozkład średniej w próbie zbliża się do normalnego, dzięki czemu możliwe było testowanie pewnych hipotez przy pomocy standardowych metod.

Wyniki przeprowadzonych testów nie dają podstaw do stwierdzenia, że któraś ze zmiennych posiada zdolności poprawnego rozróżniania do trzech grup. Wskazują na to niskie wyniki VUS (około 30%) i rozszerzonego indeksu Youdena (około 16%). Najlepsze wyniki są otrzymywane dla zmiennej BPRS.Depression, chociaż i dla niej prawdopodobieństwa dobrego przypisania do każdej z grup są rzędu 50%.

Podczas analizy dostrzeżony został błąd w funkcji **Youden3Grp**, ponieważ dopuszcza ona możliwość, że prawdopodobieństwo ma wartość ujemną. Problem ten wart jest uwagi.





## Dodatek A

# Dowód wieloczynnikowej metody delta

Tutaj znajduje się dowód twierdzenia 1.3.1.

**Twierdzenie A.0.1 (Wieloczynnikowa metoda delta)** Niech  $X^{(1)}, \dots, X^{(n)}$  będzie próbką losową o wartości oczekiwanej  $\mathbf{E}(X^{(i)}) = \mu$  i kowariancji  $\mathbf{E}(X^{(i)} - \mu)(X^{(i)} - \mu)^T = \Sigma$ . Dla danej funkcji  $g$ , która ma ciągłą pierwszą cząstkową pochodną, zachodzi

$$\sqrt{n} \left( g(\hat{X}) - g(\mu) \right) \rightarrow \mathcal{N} \left( 0, \nabla^T g(\mu) \Sigma \nabla g(\mu) \right) \text{ według rozkładu.}$$

*Dowód:* Korzystając z rozwinięcia Taylora w punkcie  $\mu$  do dwóch pierwszych wyrazów:

$$g(\hat{X}) \approx g(\mu) + \nabla g(\mu)^T (\hat{X} - \mu).$$

Z tego wynika, że wariancja  $g(\hat{X})$  w przybliżeniu wynosi:

$$\begin{aligned} \text{Var} \left( g(\hat{X}) \right) &\approx \text{Var} \left( g(\mu) + \nabla g(\mu)^T \cdot (\hat{X} - \mu) \right) = \text{Var} \left( g(\mu) + \nabla g(\mu)^T \cdot \hat{X} - \nabla g(\mu)^T \cdot \mu \right) = \\ &\text{Var} \left( \nabla g(\mu)^T \cdot \hat{X} \right) = \nabla g(\mu)^T \cdot \text{Var}(\hat{X}) \cdot \nabla g(\mu) = \nabla g(\mu)^T \cdot (\Sigma/n) \cdot \nabla g(\mu). \end{aligned}$$

Zatem otrzymujemy:

$$\left( g(\hat{X}) - g(\mu) \right) \rightarrow \mathcal{N} \left( 0, \nabla^T g(\mu) \cdot \Sigma/n \cdot \nabla g(\mu) \right)$$

i mnożąc przez  $\sqrt{n}$ :

$$\sqrt{n} \left( g(\hat{X}) - g(\mu) \right) \rightarrow \mathcal{N} \left( 0, \nabla^T g(\mu) \cdot \Sigma \cdot \nabla g(\mu) \right).$$



## Dodatek B

### Częściowe pochodne $t_-^*$ , $t_-^*$ i $J$

Częściowe pochodne  $t_-^*$ ,  $t_-^*$  i  $J$  po  $\mu_i$  i  $\sigma_i$  dla  $i=+, \theta, -$ , (z [11]).

$$\frac{\partial t_-^*}{\partial \mu_-} = -\frac{\sigma_0}{\sigma_-^2 - \sigma_0^2} \left[ \sigma_0 + \frac{\sigma_-}{\sqrt{((\mu_- - \mu_0)^2 + (\sigma_-^2 - \sigma_0^2) \ln\left(\frac{\sigma_-^2}{\sigma_0^2}\right))}} (\mu_- - \mu_0) \right], \quad (\text{B.1})$$

$$\frac{\partial t_-^*}{\partial \mu_0} = -\frac{\sigma_-}{\sigma_-^2 - \sigma_0^2} \left[ \sigma_- + \frac{\sigma_0}{\sqrt{((\mu_- - \mu_0)^2 + (\sigma_-^2 - \sigma_0^2) \ln\left(\frac{\sigma_-^2}{\sigma_0^2}\right))}} (\mu_- - \mu_0) \right], \quad (\text{B.2})$$

$$\begin{aligned} \frac{\partial t_-^*}{\partial \sigma_-} = & \frac{\left( 2\mu_0\sigma_- - \sigma_0\sqrt{((\mu_- - \mu_0)^2 + (\sigma_-^2 - \sigma_0^2) \ln\left(\frac{\sigma_-^2}{\sigma_0^2}\right))} \right) -}{(\sigma_-^2 - \sigma_0^2)^2} \\ & - \frac{\frac{\sigma_0}{\sqrt{((\mu_- - \mu_0)^2 + (\sigma_-^2 - \sigma_0^2) \ln\left(\frac{\sigma_-^2}{\sigma_0^2}\right))}} \left[ \sigma_-^2 \ln\left(\frac{\sigma_-^2}{\sigma_0^2}\right) + (\sigma_-^2 - \sigma_0^2) \right] (\sigma_-^2 - \sigma_0^2) -}{(\sigma_-^2 - \sigma_0^2)^2} \\ & - \frac{2\sigma_- \left( (\mu_0\sigma_-^2 - \mu_- \sigma_0^2) - \sigma_- \sigma_0 \sqrt{((\mu_- - \mu_0)^2 + (\sigma_-^2 - \sigma_0^2) \ln\left(\frac{\sigma_-^2}{\sigma_0^2}\right))} \right)}{(\sigma_-^2 - \sigma_0^2)^2}, \quad (\text{B.3}) \end{aligned}$$

$$\begin{aligned}
\frac{\partial t_-^*}{\partial \sigma_0} = & \frac{\left(-2\mu_- \sigma_0 - \sigma_- \sqrt{((\mu_- - \mu_0)^2 + (\sigma_-^2 - \sigma_0^2) \ln(\frac{\sigma_-^2}{\sigma_0^2}))}\right) +}{(\sigma_-^2 - \sigma_0^2)^2} \\
& + \frac{\sigma_-}{\sqrt{((\mu_- - \mu_0)^2 + (\sigma_-^2 - \sigma_0^2) \ln(\frac{\sigma_-^2}{\sigma_0^2}))}} \left[ \sigma_0^2 \ln\left(\frac{\sigma_-^2}{\sigma_0^2}\right) + (\sigma_-^2 - \sigma_0^2) \right] (\sigma_-^2 - \sigma_0^2) + \\
& \frac{+ 2\sigma_0 \left( (\mu_0 \sigma_-^2 - \mu_- \sigma_0^2) - \sigma_- \sigma_0 \sqrt{((\mu_- - \mu_0)^2 + (\sigma_-^2 - \sigma_0^2) \ln(\frac{\sigma_-^2}{\sigma_0^2}))} \right)}{(\sigma_-^2 - \sigma_0^2)^2}. \quad (B.4)
\end{aligned}$$

Ponieważ  $t_+^*$  ma analogiczny wzór do  $t_-^*$ , odpowiednie pochodne po  $\mu_0$ ,  $\mu_+$ ,  $\sigma_0$  oraz  $\sigma_+$  można zapisać zastępując w powyższych wzorach:  $\mu_-$  na  $\mu_0$ ,  $\mu_0$  na  $\mu_+$ ,  $\sigma_-$  na  $\sigma_0$  oraz  $\sigma_0$  na  $\sigma_+$ .

Pochodne cząstkowe indeksu Youdena po kolejnych zmiennych są następujące:

$$\frac{\partial J}{\partial \mu_-} = \frac{1}{2} \left\{ \frac{\partial t_-^*}{\partial \mu_-} \left[ \frac{1}{\sigma_-} \varphi\left(\frac{t_- - \mu_-}{\sigma_-}\right) - \frac{1}{\sigma_0} \varphi\left(\frac{t_- - \mu_0}{\sigma_0}\right) \right] - \frac{1}{\sigma_-} \varphi\left(\frac{t_- - \mu_-}{\sigma_-}\right) \right\}, \quad (B.5)$$

$$\frac{\partial J}{\partial \mu_+} = -\frac{1}{2} \left\{ \frac{\partial t_+^*}{\partial \mu_+} \left[ \frac{1}{\sigma_+} \varphi\left(\frac{t_+ - \mu_+}{\sigma_+}\right) - \frac{1}{\sigma_0} \varphi\left(\frac{t_+ - \mu_0}{\sigma_0}\right) \right] - \frac{1}{\sigma_+} \varphi\left(\frac{t_+ - \mu_+}{\sigma_+}\right) \right\}, \quad (B.6)$$

$$\begin{aligned}
\frac{\partial J}{\partial \mu_0} = & \frac{1}{2} \left\{ \frac{\partial t_-^*}{\partial \mu_0} \left[ \frac{1}{\sigma_-} \varphi\left(\frac{t_- - \mu_-}{\sigma_-}\right) - \frac{1}{\sigma_0} \varphi\left(\frac{t_- - \mu_0}{\sigma_0}\right) \right] + \frac{\partial t_+^*}{\partial \mu_0} \left[ \frac{1}{\sigma_0} \varphi\left(\frac{t_+ - \mu_0}{\sigma_0}\right) - \right. \right. \\
& \left. \left. - \frac{1}{\sigma_+} \varphi\left(\frac{t_+ - \mu_+}{\sigma_+}\right) \right] + \frac{1}{\sigma_0} \left[ \varphi\left(\frac{t_- - \mu_0}{\sigma_0}\right) - \varphi\left(\frac{t_+ - \mu_0}{\sigma_0}\right) \right] \right\}, \quad (B.7)
\end{aligned}$$

$$\frac{\partial J}{\partial \sigma_-} = \frac{1}{2} \left\{ \frac{\partial t_-^*}{\partial \sigma_-} \left[ \frac{1}{\sigma_-} \varphi\left(\frac{t_- - \mu_-}{\sigma_-}\right) - \frac{1}{\sigma_0} \varphi\left(\frac{t_- - \mu_0}{\sigma_0}\right) \right] - \frac{t_- - \mu_-}{\sigma_-^2} \varphi\left(\frac{t_- - \mu_-}{\sigma_-}\right) \right\}, \quad (B.8)$$

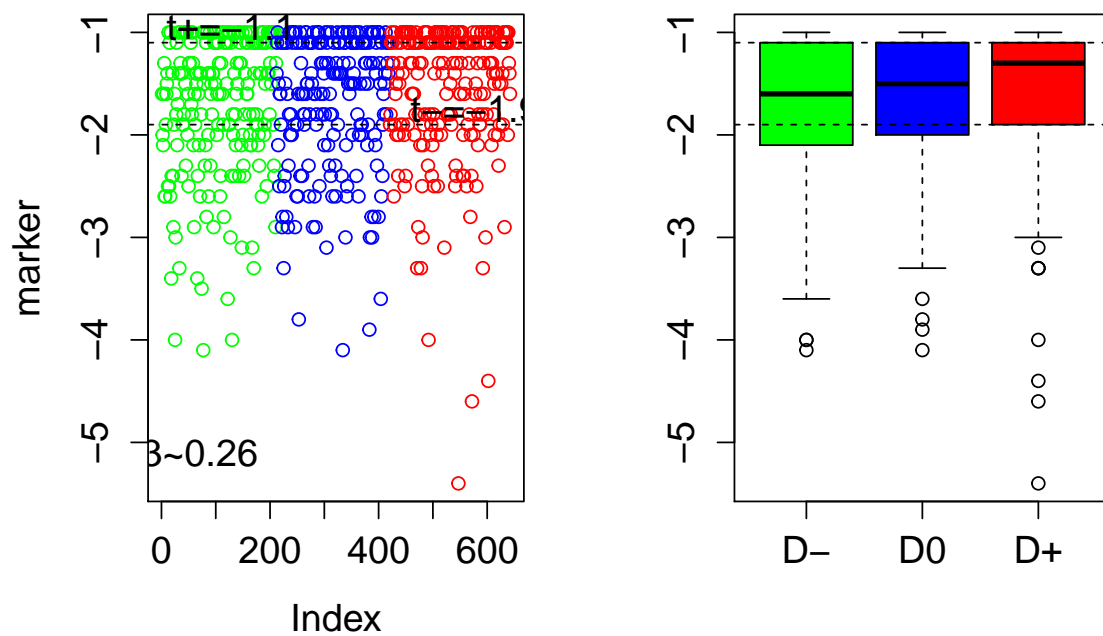
$$\frac{\partial J}{\partial \sigma_+} = -\frac{1}{2} \left\{ \frac{\partial t_+^*}{\partial \sigma_+} \left[ \frac{1}{\sigma_+} \varphi\left(\frac{t_+ - \mu_+}{\sigma_+}\right) - \frac{1}{\sigma_0} \varphi\left(\frac{t_+ - \mu_0}{\sigma_0}\right) \right] - \frac{t_+ - \mu_+}{\sigma_+^2} \varphi\left(\frac{t_+ - \mu_+}{\sigma_+}\right) \right\}, \quad (B.9)$$

$$\begin{aligned}
\frac{\partial J}{\partial \sigma_0} = & \frac{1}{2} \left\{ \frac{\partial t_-^*}{\partial \sigma_0} \left[ \frac{1}{\sigma_-} \varphi\left(\frac{t_- - \mu_-}{\sigma_-}\right) - \frac{1}{\sigma_0} \varphi\left(\frac{t_- - \mu_0}{\sigma_0}\right) \right] + \frac{\partial t_+^*}{\partial \sigma_0} \left[ \frac{1}{\sigma_0} \varphi\left(\frac{t_+ - \mu_0}{\sigma_0}\right) - \right. \right. \\
& \left. \left. - \frac{1}{\sigma_+} \varphi\left(\frac{t_+ - \mu_+}{\sigma_+}\right) \right] + \frac{t_- - \mu_0}{\sigma_0^2} \varphi\left(\frac{t_- - \mu_0}{\sigma_0}\right) - \frac{t_+ - \mu_0}{\sigma_0^2} \varphi\left(\frac{t_+ - \mu_0}{\sigma_0}\right) \right\}. \quad (B.10)
\end{aligned}$$

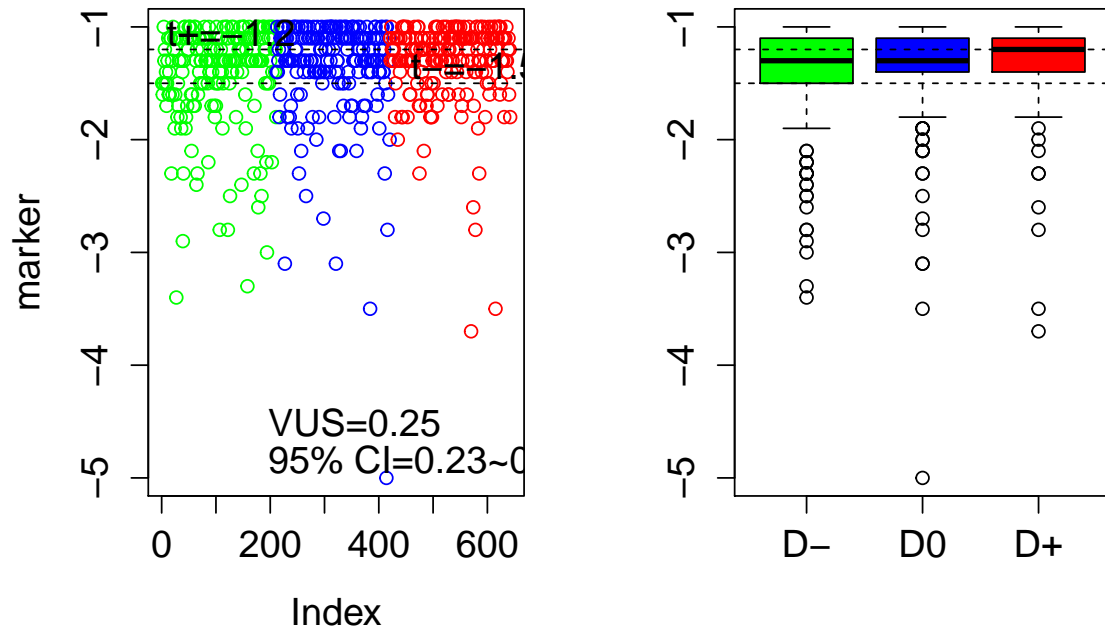
## Dodatek C

# Wykresy pudełkowe i rozproszenia VUS

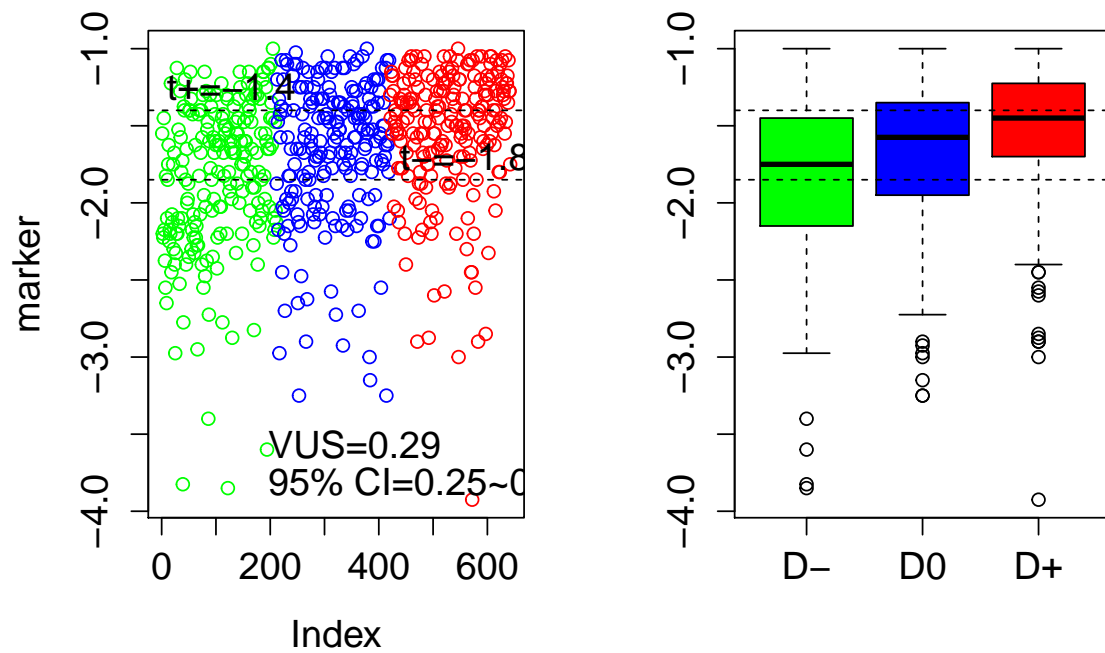
Rysunek C.1: Wykres rozproszenia oraz wykres pudełkowy BPRS.Negative



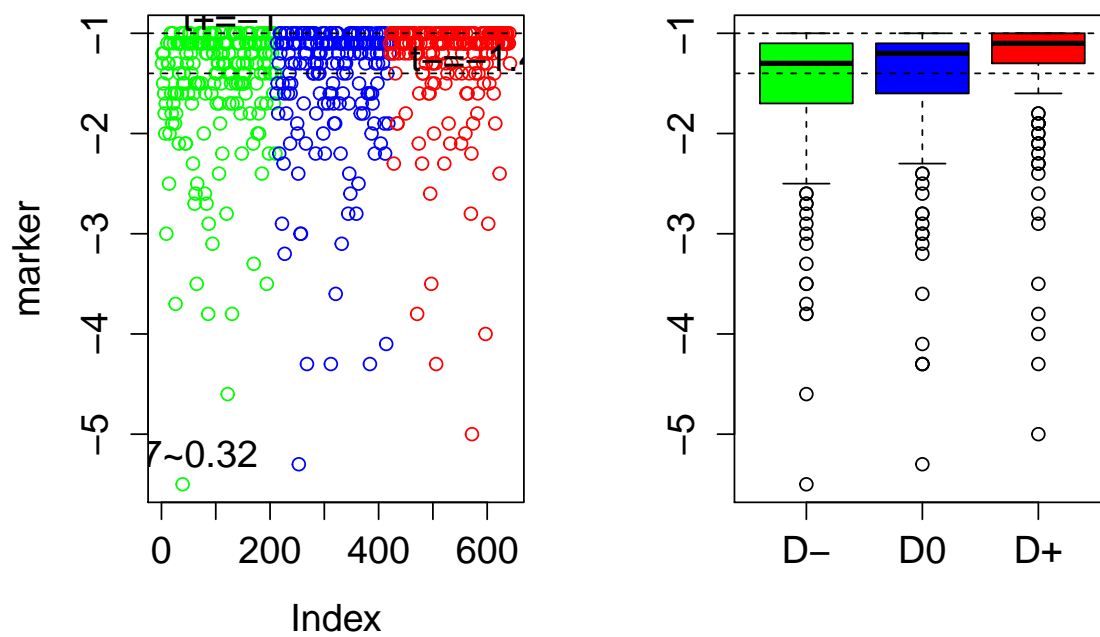
**Rysunek C.2:** Wykres rozproszenia oraz wykres pudełkowy BPRS.Average



**Rysunek C.3:** Wykres rozproszenia oraz wykres pudełkowy BPRS.Maniac



Rysunek C.4: Wykres rozproszenia oraz wykres pudełkowy BPRS.Positive







# Bibliografia

- [1] Adamowski Tomasz, Hadryś Tomasz, Kiejna Andrzej, *Efektywność leczenia na psychiatrycznym oddziale dziennym, w porównaniu z oddziałem stacjonarnym, na podstawie analizy objawów psychopatologicznych, subiektywnej oceny jakości życia oraz częstości rehospitalizacji po zakończeniu leczenia*, Psychiatria Polska 2008, tom XLII, numer 4, 571 — 574.
- [2] Akobeng Anthony, *Understanding diagnostic tests3: receiver operating characteristic curves*, Acta paediatrica 2007, May 96(5), 644-646.
- [3] Biecek Przemysław, *Package „PBImisc”*, <http://cran.r-project.org/web/packages/PBImisc/>, [18.03.2013].
- [4] Casella George, Berger Roger L. *Statistical Inference*, Duxbury Advanced Series, 2002, 245.
- [5] Ferri C., Hernandez-Orallo J., M.A. Salido, *Volume under the ROC surface for multi-class problems*, Lecture Notes in Computer Science, 108.
- [6] Hand David J., Till Robert J., *A simple generalisation of the area under the ROC curve for multiple class classification problems*, Machine Learning, Volume 45, 2001, 171 — 186.
- [7] Hanley James, McNeil Barbara, *The meaning and use of the area under Receiver operating characteristic (ROC) curve*, Radiology, Volume 143, Issue 1, 2010, 29 — 36.
- [8] Harańczyk Grzegorz, *Krzywe ROC, czyli ocena jakości klasyfikatora i poszukiwanie optymalnego punktu odcięcia*, StatSoft Polska, 2010.
- [9] Jialiang Li, *ROC analysis with multiple classes and multiple tests: methodology and its application in microarray studies*, Biostatistics, Volume 9, Issue 3, 2008, 569.
- [10] Jingqin Luo, Chengjie Xiong, *DiagTest3Grp: An R package for analyzing diagnostic tests with three ordinal groups*, Journal of Statistical Software, Volume 51, Issue 3, 2012.
- [11] Jingqin Luo, Chengjie Xiong, *Youden Index and Associated Cut-Points for Three Ordinal Diagnostic Groups*, Communications in Statistics - Simulation and Computation, 42:6, 1213 — 1234.
- [12] Luo Jingqin, Xiong Chengjie, *DiagTest3Grp: Diagnostic test summary measures for three ordinal groups*, <http://cran.r-project.org/web/packages/DiagTest3Grp/index.html>, [11.05.2013].

- [13] Luo Jingqin , Xiong Chengjie, *Package DiagTest3Grp*, <http://cran.r-project.org/web/packages/DiagTest3Grp/DiagTest3Grp.pdf>, [11.05.2013].
- [14] Lumley Thomas, Diehr Paula, Scott Emerson, Lu Chen, *The importance of the normality assumption in large public health data sets*, Annu Rev Public Health 2002, 23, 151-169.
- [15] Niemiro Wojciech, *Skrypt do przedmiotu Statystyka*, [http://www.mimuw.edu.pl/~wniem/Statystyka\\_I/Statystyka.pdf](http://www.mimuw.edu.pl/~wniem/Statystyka_I/Statystyka.pdf) [06.04.2013]
- [16] Ripley Brian, *Package „MASS"*, <http://cran.r-project.org/web/packages/MASS/>, [29.05.2013].
- [17] Shapiro Samuel Stanford, Wilk Martin Bradbury, *An analysis of variance test for normality (complete samples*, Biometrika, Volume 52, No. 3/4, 1965, 591 — 611.
- [18] Sheather Simon J., Jones M.C., *A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation*, Journal of Royal Statistical Society B, Volume 53, 1991, 683 — 690.
- [19] Sheather Simon J., *The Performance of Six Popular Bandwidth Selection Methods on Some Real Datasets*, Computational Statistics, Volume 7, 1992, 225 — 250.
- [20] Sing Tobias, Sander Oliver, *ROCR: Visualizing the performance of scoring classifiers*, <http://cran.r-project.org/web/packages/ROCR/index.html>, [20.06.2013]
- [21] Xiong Chengjie, Gerald van Belle, *A parametric comparison of diagnostic accuracy with three ordinal diagnostic groups*, Biometrical Journal 2007, Volume 5, 684-686.
- [22] Xiong Chengjie, Gerald van Belle, *Measuring and estimating diagnostic accuracy when there are three ordinal diagnostic groups*, Statistics in medicine 2006, 25.
- [23] *Statistics with R: Box-Cox transformation of the response in a linear model* [online], dostępny w Internecie: <http://www.youtube.com/attribution?v=TgVx9Rqsewo> [07.05.2013].