

# STATISTICA Zestaw Medyczny 2.0

## Spis treści

<b>1. Czyszczenie danych</b> .....	<b>2</b>
1.1. Poprawność danych.....	2
1.2. Analiza brakujących danych .....	2
1.3. Zmienne sztuczne.....	3
<b>2. Analizy</b> .....	<b>4</b>
2.1. Krzywe ROC (Receiver Operating Characteristic) .....	4
2.2. Metaanaliza i metaregresja.....	4
2.3. Kreator regresji logistycznej .....	5
<b>3. Analizy dodatkowe</b> .....	<b>7</b>
3.1. Miary powiązania/efektów.....	7
3.2. Wykres Blanda-Altmana.....	7

# Zestaw Medyczny STATISTICA

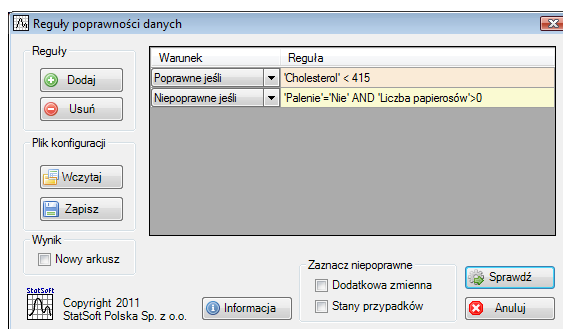
**STATISTICA Zestaw Medyczny** jest specjalistycznym pakietem analitycznym, który stanowi uzupełnienie funkcjonalności programów z rodziny **STATISTICA**. W skład pakietu **STATISTICA Zestaw Medyczny** wchodzi następujące moduły:

- Czyszczenie danych
  - Poprawność danych
  - Braki danych
  - Zmienne sztuczne
- Analizy
  - Krzywe ROC
  - Metaanaliza i metaregresja
  - Kreator regresji logistycznej
- Analizy dodatkowe
  - Miary powiązania/efektów
  - Wykres Blanda-Altmana
  - Test post hoc ANOVA Friedmana

## 1. Czyszczenie danych

### 1.1. Poprawność danych

Dzięki tej opcji użytkownik ma możliwość łatwiejszego zdefiniowania reguł poprawności danych. W oknie definiowania reguł można wybrać nazwę zmiennej, wskazać żądane wartości lub odpowiadające im etykiety i połączyć je odpowiednim operatorem. Oprócz prostych reguł logicznych użytkownik ma możliwość wykorzystania zaawansowanej biblioteki funkcji zawartej w **STATISTICA**. Generowanie reguł ułatwia specjalnie przygotowany kreator. W arkuszu utworzone zostają zmienne wskazujące przypadki poprawne z punktu widzenia określonej reguły. Dodatkowo można utworzyć zmienną sprawdzającą poprawność względem wszystkich podanych reguł. Przygotowane reguły możemy zapisać do pliku konfiguracyjnego i wykorzystywać w innych modułach programu.

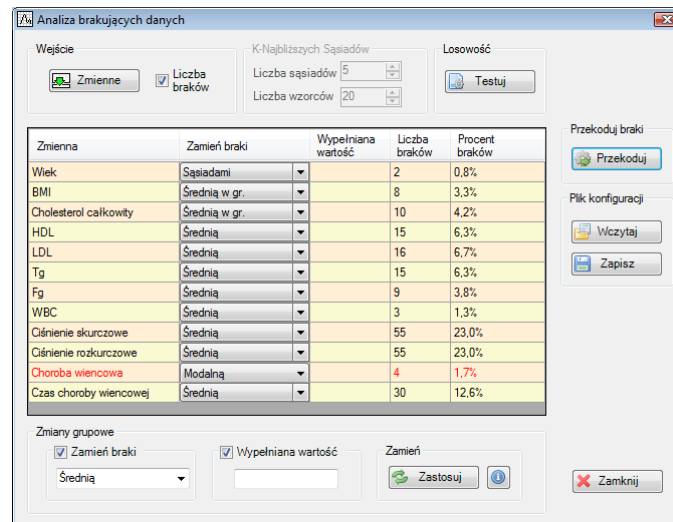


### 1.2. Analiza brakujących danych

Moduł umożliwia przekodowanie braków danych według wskazanego schematu

- Bogaty zestaw sposobów imputacji braków danych

- Średnią, medianą, modalną
- Średnią bądź medianą w grupach
- Najbliższymi sąsiadami
- Podaną wartością
- Łatwe określanie tej samej akcji dla wielu zmiennych
- Testowanie losowości braków danych
- Zapis określonych schematów kodowania do pliku konfiguracji

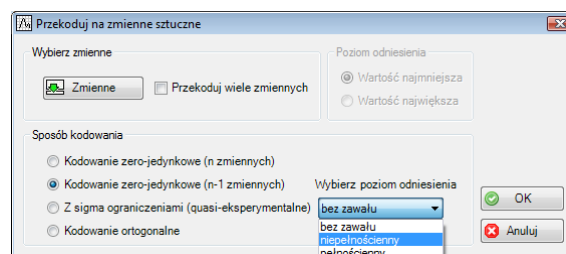


### 1.3. Zmienne sztuczne

Moduł umożliwia zamianę cech jakościowych na odpowiadające im zmienne sztuczne. W programie zaimplementowano cztery schematy kodowania:

- Kodowanie zero-jedynkowe na n zmiennych (n to liczba poziomów kodowanej cechy),
- Kodowanie zero-jedynkowe na n-1 zmiennych,
- Kodowanie z sigma ograniczeniami (quasi-eksperymentalne),
- Kodowanie ortogonalne.

W przypadku wyboru trzech ostatnich schematów mamy możliwość ręcznego wyboru poziomu odniesienia.



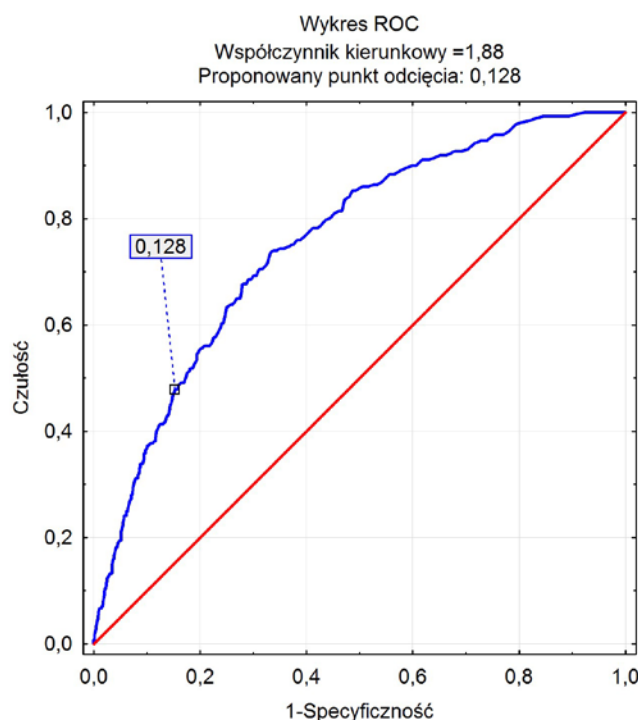
## 2. Analizy

### 2.1. Krzywe ROC (Receiver Operating Characteristic)

Krzywe ROC są narzędziem służącym do oceny poprawności klasyfikatora (pojedynczej zmiennej lub całego modelu), zapewniają one łączny opis jego czułości i specyficzności. Ten sposób wspomagania systemu decyzyjnego jest szeroko stosowany w różnych obszarach analizy danych, m.in. w diagnostyce medycznej.

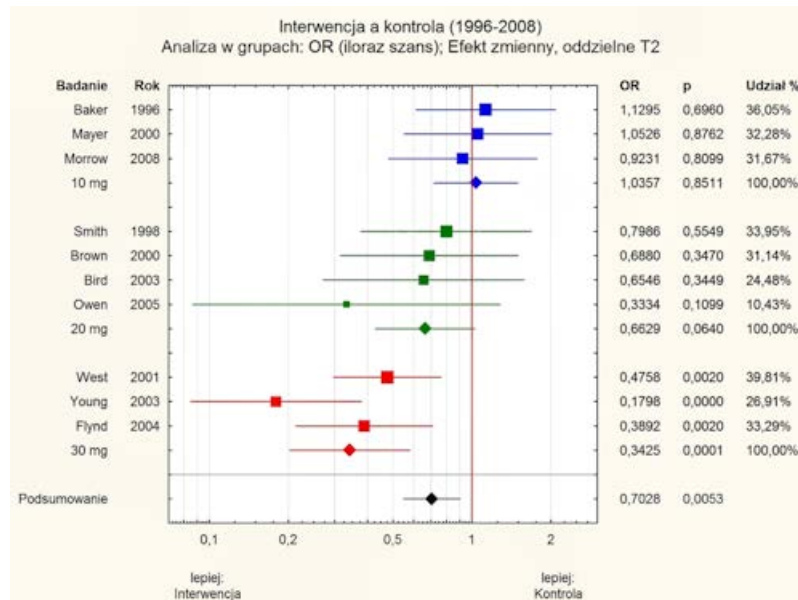
Moduł *Krzywe ROC* umożliwia:

- kreślenie krzywych ROC dla prób zależnych i niezależnych,
- obliczanie pola powierzchni pod krzywą,
- porównywanie istotności różnicy pól pomiędzy dwiema krzywymi
- ustalanie optymalnego punktu odcięcia dla podanych kosztów błędnej klasyfikacji i prawdopodobieństw a priori występowania badanego zjawiska,
- obliczanie miar FP, TP, FN, FP, Sensitivity, Specificity, ACC, PPV, NPV, False positive ratio, False negative ratio, LR dla wszystkich możliwych punktów odcięcia
- wykresy czułości i specyficzności.



### 2.2. Metaanaliza i metaregresja

Moduł *Metaanaliza i metaregresja* jest narzędziem umożliwiającym syntezę wyników wielu niezależnych badań szczególnie w sytuacji, gdy nie mamy dostępu do danych surowych a dysponujemy jedynie zbiorczymi wynikami tych badań. Podejście takie pozwala rozszerzyć wnioski z pojedynczych badań na szerszą populację oraz zwiększyć wiarygodność otrzymanych wyników. Za pomocą dodatkowych narzędzi takich jak analiza niejednorodności, analiza w grupach czy metaregresja badacz może również ocenić zmienność uzyskanych wyników i wskazać jej źródła. Obliczenia są wykonywane dla szeregu miar efektu, a wyniki można przedstawić w postaci szczegółowych raportów oraz wykresów.



Moduł *Metaanaliza i metaregresja* umożliwia m.in.:

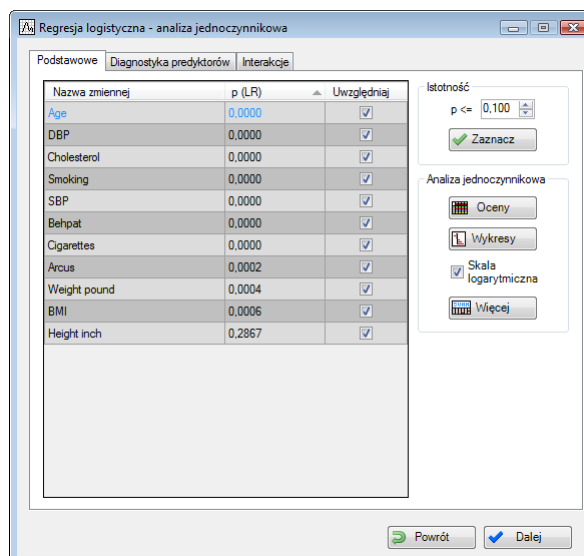
- wprowadzenie wyników badań i gotowych (już wyliczonych) miar efektu;
- wprowadzanie wyników przedstawionych w różnych formatach;
- uwzględnienie w analizie badań z wynikami przedstawionymi w postaci różnych miar efektu (np. ilorazu szans oraz d Cohena);
- obliczanie miar efektu dla pojedynczych badań;
- obliczanie łącznych miar efektu dla modelu z efektem stałym i zmiennym (*fixed effect model, random effects model*);
- wykonanie metaanalizy (*meta-analysis*) i utworzenie wykresu leśnego (*forest plot*);
- przeprowadzenie analizy skumulowanej (*cumulative meta-analysis*);
- wykonanie analizy niejednorodności (*heterogeneity analysis*) - miary  $Q$ ,  $T^2$ ,  $I^2$ ;
- utworzenie wykresów: L'Abbego i Galbraitha;
- przeprowadzenie metaanalizy w grupach (*subgroup analyses*) dla modelu z efektem stałym oraz efektem zmiennym, z oddzielnym  $T^2$  i wspólnym  $T^2$ ;
- wykonanie metaregresji (*meta-regression*) i przedstawienie jej wyników w postaci raportów i wykresu bąblowego;
- przeprowadzenie analizy wrażliwości (*sensitivity analysis*) - analiza po dołączeniu grupy badań lub wyłączeniu wybranej kombinacji badań;

### 2.3. Kreator regresji logistycznej

Moduł umożliwia zbudowanie i ocenę modelu regresji logistycznej. Korzystając z Kreatorabadaacz krok po kroku wykonuje kolejne etapy związane z budową modelu regresji poczynając od sposobu kodowania zmiennych oraz wyboru istotnych cech do analizy, poprzez sprawdzanie założeń i identyfikację interakcji aż po ocenę dobroci dopasowania modelu, analizę reszt czy zbadanie jego zdolności predykcyjnych. Funkcjonalność Kreatora obejmuje między innymi:

- Wygodne określenie modelowanej klasy zmiennej zależnej oraz poziomów odniesienia predyktorów jakościowych
- Wykonanie jednoczynnikowej analizy dla wszystkich wybranych predyktorów
  - Oceny parametrów regresji
  - Obliczanie ilorazów szans wraz z przedziałami ufności

- Wykres leśny (*forest plot*) dla zmiennych jakościowych
- Ranking istotności predyktorów na podstawie testu LR



- Badanie liniowości wpływu predyktorów ilościowych na logarytm szansy modelowanego zjawiska
- Analiza współliniowości predyktorów
  - Analiza korelacji
  - Wyznaczanie skupisk skorelowanych zmiennych
  - Automatyczne wyznaczenie reprezentantów zidentyfikowanych skupisk zmiennych
- Analiza interakcji
  - Automatyczna identyfikacja istotnych interakcji
  - Tworzenie rankingu interakcji
  - Wygodny wybór interesujących interakcji
- Krokowe metody doboru zmiennych do modelu
  - Krokowa postępująca i wsteczna
  - Wprowadzanie postępujące
  - Eliminacja wsteczna
- Zaawansowane schematy walidacji modelu
  - Próba ucząca i testowa
  - Wielokrotna ocena krzyżowa
  - Metoda LOO (*Leave One Out*)
- Bogaty zestaw miar jakości modelu:
  - Testy LR
  - Odchylenie
  - Miary pseudo  $R^2$
  - AIC, BIC
  - Test Hosmera-Lemeshowa
- Analiza reszt oraz wartości wpływowych
- Analiza krzywych ROC
- Wykresy przyrostu i zysku

### 3. Analizy dodatkowe

#### 3.1. Miary powiązania/efektów

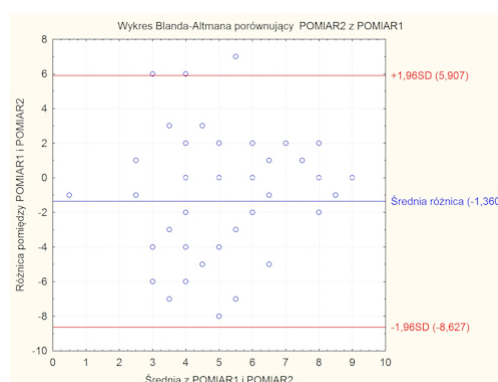
Miary powiązania/efektów jest dedykowanym modułem przeznaczonym do obliczania na podstawie tabeli 2x2 szeregu wskaźników powiązania lub efektu. Na przykład umożliwia on obliczenie efektu związanego z binarną zmienną zależną, spowodowanego manipulacją zmienną niezależną. Moduł umożliwia zarówno obliczenie wskaźników na podstawie danych surowych, jak również ręczne wprowadzenie lub korektę wartości w tabeli.

Miary powiązania - Miary efektu	
Ryzyko eksponowanych - DWP (PPV)	0.600
Ryzyko nieeksponowanych - UWP (NPV)	0.250
Ryzyko populacji	0.444
Bezwzględne zmniejszenie ryzyka (ARR)	0.350
Liczba wymaganych (NNT)	2.857
Ryzyko względne (RR)	2.400
Szansa stanu "Tak" (ODDS)	1.500
Szansa stanu "Nie" (ODDS)	3.000
Iloraz szans (OR)	4.500
Dokładność (ACC)	0.667
Czułość (Sensitivity)	0.750
Specyficzność (Specificity)	0.600
Dodatni iloraz wiarygodności (LR+)	1.875
Ujemny iloraz wiarygodności (LR-)	0.417
Wskaźnik J Yourdena	0.350

Miary te mają szczególne znaczenie w diagnostycznej analizie związków przyczynowych na podstawie rozkładów w tabeli obrazującej relacje typu test-efekt.

#### 3.2. Wykres Blanda-Altmana

Wykres Blanda-Altmana wykorzystywany najczęściej w chemometrii i biostatystyce przedstawia stopień zgodności pomiędzy dwiema różnymi próbkami bądź wskaźnikami. Merytorycznie jest on identyczny z wykresem średnia-różnica Tukeya.



Jednym z głównych zastosowań wykresu Blanda-Altmana jest porównanie dwóch wskaźników klinicznych, z których każdy obciążony jest pewnym błędem pomiaru. Może być on także wykorzystany do porównania nowej techniki pomiaru bądź wskaźnika z obowiązującym złotym standardem.

Na wykresie przedstawia się przedziały zgodności liczone jako średnia różnica pomiędzy badanymi wskaźnikami  $\pm 1,96^*$  odchylenie standardowe różnicy.