

# STATISTICA 10

## Nowości i udoskonalenia

### Spis treści

<b>1. Wydajność .....</b>	<b>2</b>
<b>2. Współpraca z innymi systemami .....</b>	<b>2</b>
2.1. SharePoint.....	2
2.2. Office 2010.....	2
2.3. OLAP.....	2
2.4. STATISTICA PI Connector.....	3
<b>3. Wykresy .....</b>	<b>3</b>
3.1. Opis ogólny .....	3
3.2. Skalowanie interakcyjne .....	4
3.3. Panoramowanie interakcyjne .....	4
3.4. Przezroczystość .....	4
3.5. Linie odniesienia.....	5
3.6. Bezpośrednia edycja tekstów.....	5
<b>4. Środowisko pracy, ergonomia.....</b>	<b>5</b>
4.1. Opis ogólny .....	5
4.2. Wstążka .....	6
4.3. Przestrzeń robocza data mining .....	6
4.4. Dodatki STATISTICA na wstążce.....	7
<b>5. Statystyki.....</b>	<b>7</b>
5.1. Symulacja rozkładów i macierzy korelacji .....	7
5.2. Modele proporcjonalnego hazardu Coxa .....	8
5.3. Dalsze udoskonalenia wydajności i skalowalności w obliczaniu statystyk opisowych .....	8
<b>6. Data Mining .....</b>	<b>9</b>
6.1. Wykres zysku.....	9
6.2. Krzywa ROC.....	9
6.3. Dalsze udoskonalenia wydajności i skalowalności metod predykcyjnych .....	9
6.4. Text Mining .....	9
6.5. Wdrożenie w Java i C# .....	10
6.6. STATISTICA Zestaw Skoringowy .....	10
<b>7. STATISTICA Enterprise .....</b>	<b>11</b>
7.1. Wstążka .....	11
7.2. Konfiguracje danych .....	11
7.3. Przenoszenie bazy danych .....	11
7.4. Wdrażanie makr w STATISTICA Enterprise .....	12
7.5. Nazwy konfiguracji .....	12
7.6. STATISTICA Enterprise Server - automatyczna aktualizacja kart kontrolnych .....	12
7.7. STATISTICA Enterprise Server - eksploracja karty kontrolnej.....	12
<b>8. STATISTICA MSPC Online.....</b>	<b>13</b>
<b>9. STATISTICA Web Data Entry .....</b>	<b>13</b>
<b>10. Dedykowane rozwiązania skoringowe .....</b>	<b>13</b>
10.1. STATISTICA Zestaw Skoringowy .....	13
10.2. STATISTICA Live Score .....	14
<b>12. Pomoc i dokumentacja .....</b>	<b>14</b>
Przykłady modelu obiektowego STATISTICA .....	14

# STATISTICA 10

## 1. Wydajność

(Wszystkie produkty)

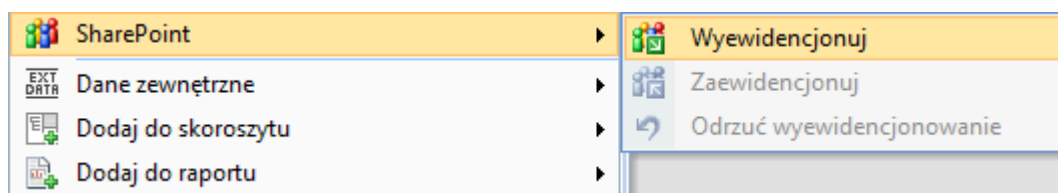
STATISTICA 10 cechuje się jeszcze większą wydajnością niż wcześniejsze wersje, osiągniętą dzięki automatycznemu wykorzystaniu 64-bitowych procesorów (jeśli tylko umożliwia to system operacyjny i sprzęt) oraz wysoce zoptymalizowanej wielowątkowości. Wiele funkcji przetwarzania danych i analiz (C&RT, CHAID, ogólne modele liniowe itd.), które w wersji 9 były jednowątkowe, teraz zostały udoskonalone, tak aby wykorzystywały wiele procesorów lub rdzeni.

## 2. Współpraca z innymi systemami

### 2.1. SharePoint

(Wszystkie produkty)

STATISTICA 10 może otwierać i zapisywać dokumenty w systemie Microsoft SharePoint, który jest najszybciej rozwijającym się standardem wymiany i integracji danych.



### 2.2. Office 2010

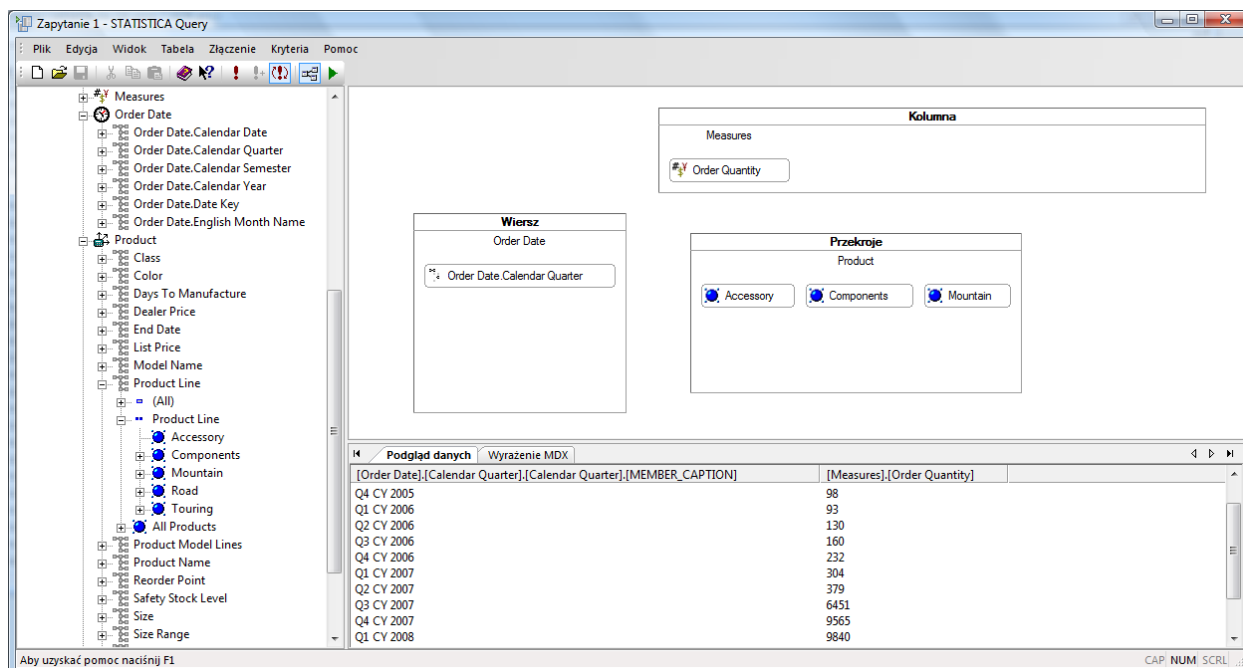
(Wszystkie produkty)

STATISTICA bezpośrednio wczytuje dane w formacie Excel 2007 i 2010 z uwzględnieniem formatowania. Nowa technologia poprawiła szybkość i niezawodność wczytywania danych z arkuszy Excel 2007 i 2010.

### 2.3. OLAP

(Wszystkie produkty)

STATISTICA Query umożliwia pobieranie danych z kostek OLAP za pomocą dostawców, takich jak Microsoft OLE DB Provider for Analysis Services lub SAP Business Warehouse. Zapytania MDX można tworzyć w środowisku graficznym lub wpisywać w trybie tekstowym (aktualnie dostępne w wersji Beta).



## 2.4. STATISTICA PI Connector

(STATISTICA PI Connector)

W STATISTICA 10 udoskonalono instalację i zarządzanie dodatkiem STATISTICA PI Connector, który jest rozprowadzany razem z resztą programu i nie jest już wymagana oddzielna instalacja.

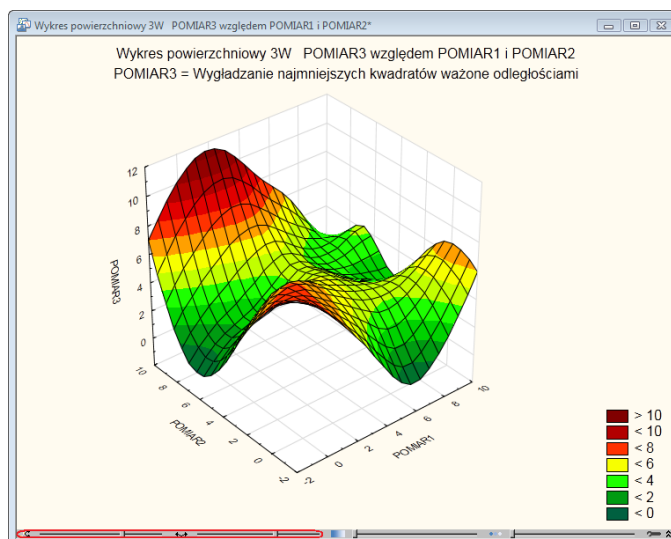
## 3. Wykresy

(Wszystkie produkty)

### 3.1. Opis ogólny

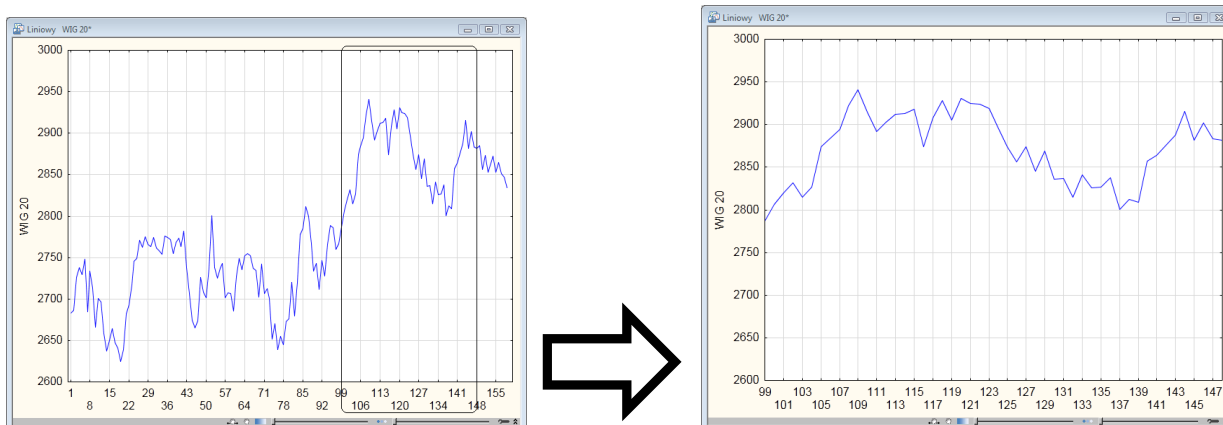
Kreślenie wykresów w STATISTICA 10 zostało znacząco udoskonalone. Program automatycznie wykrywa i wykorzystuje możliwości przyspieszania sprzętowego dostępnego w wielu, nawet podstawowych, kartach graficznych komputerów stacjonarnych i notebooków. Dzięki temu wykresy są nie tylko tworzone szybciej, ale dodatkowo dostępne są zaawansowane wygładzanie i wypełnienia gradientowe.

Na dole wszystkich okien wykresów, zarówno samodzielnych, jak i umieszczonych w skoroszybie, znajdują się elementy sterujące (kontrolki) umożliwiające określenie przezroczystości pól i znaczników na wykresie. Korzyści to nie tylko udoskonalony wygląd wykresu, ale również możliwość badania danych i wykrywania wzorców, takie jak odkrywanie zależności poprzez zmniejszanie gęstości punktów i obracanie wykresów trójwymiarowych w pionie i poziomie.



### 3.2. Skalowanie interakcyjne

W wersji 10 wprowadzono interakcyjne skalowanie osi za pomocą myszy. Wystarczy umieścić kursor myszy na wysokości wartości skali osi w jej początkowej lub końcowej części, wcisnąć lewy klawisz myszy i przesunąć kursor w kierunku środka osi, aby zakres osi został zmniejszony (tzn. aby wykres obejmował mniejszy obszar, ale z większą dokładnością). Natomiast jeśli przesuniemy kursor w kierunku końca osi (dalszego od obecnego położenia myszy), to zakres osi zostanie zwiększony (wykres będzie obejmował większy obszar, ale z mniejszą dokładnością). Takie rozciąganie i ściskanie wykresu jest użytecznym sposobem badania danych, umożliwiającym odkrycie wzorców poprzez powiększanie i pomniejszanie wybranego obszaru.

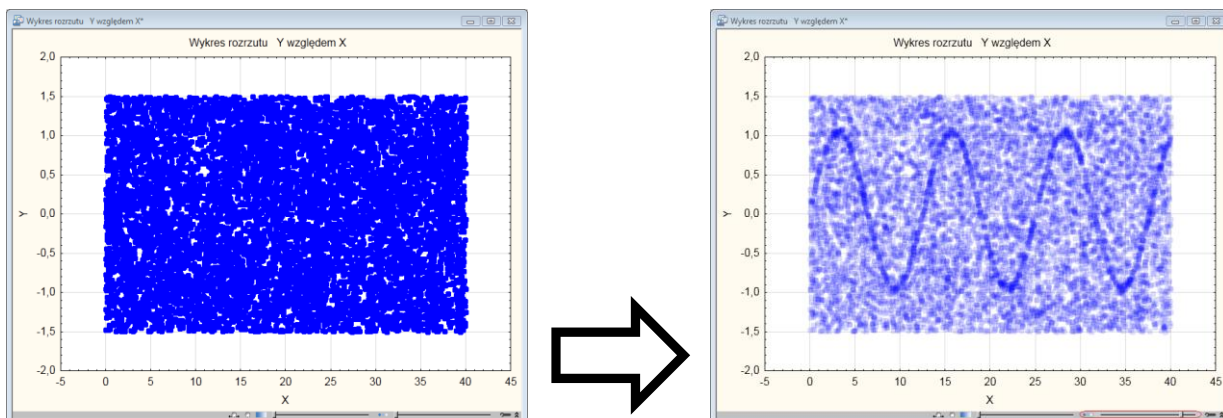


### 3.3. Panoramowanie interakcyjne

Pokazywany obszar możemy przesunąć w lewo lub w prawo. Wystarczy w tym celu umieścić kursor myszy na wysokości wartości skali osi w jej środkowej części i wcisnąć lewy klawisz myszy. Następnie przemieszczamy kursor w wybranym kierunku, aby przesunąć pokazywany obszar wykresu. Interakcyjne panoramowanie może być pomocne w wykrywaniu wzorców ukrytych w danych.

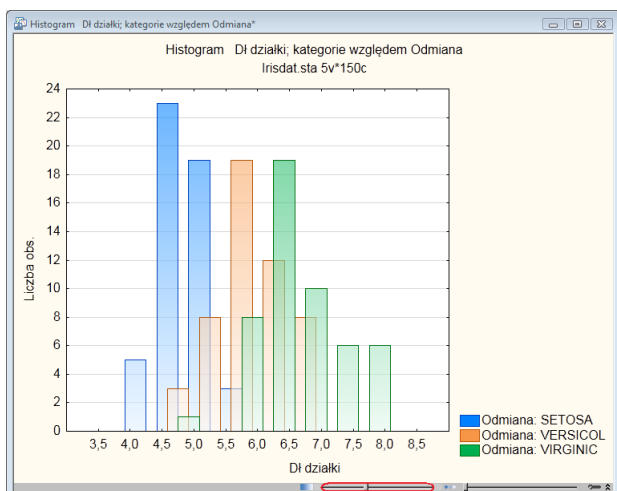
### 3.4. Przeźroczystość

W STATISTICA 10 możemy używać obszarów i znaczników o różnym stopniu przeźroczystości w środowisku Windows Vista SP 2 lub Windows 7), przy czym jest on określany interakcyjnie za pomocą suwaków umieszczonych w dolnej części wykresu. Dzięki temu bardziej widoczne są nakładające się obszary i punkty. Określanie przeźroczystości umożliwia odkrycie wzorców występujących w bardzo licznej zbiorowości punktów, zwłaszcza na wykresach rozrzutu i macierzowych utworzonych dla dużych zbiorów danych.



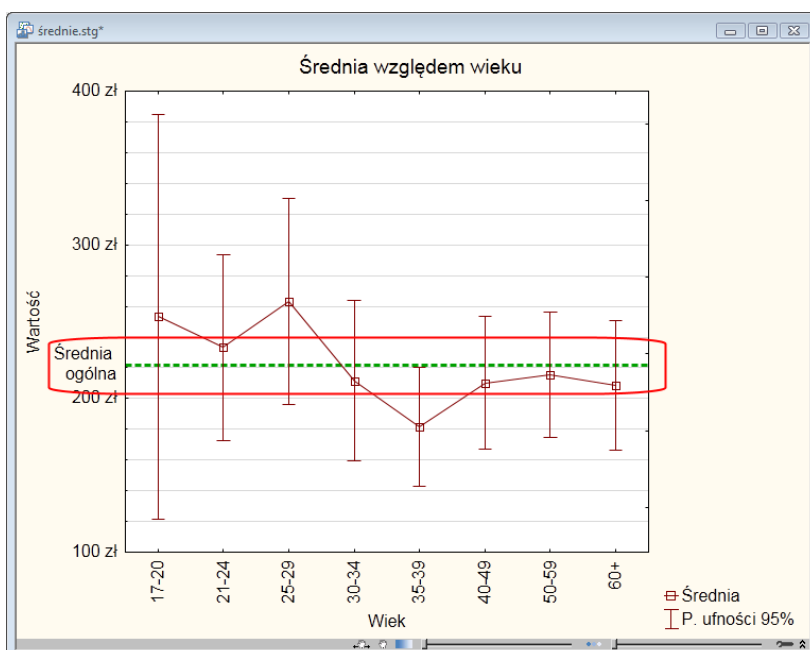
Dobranie odpowiedniej przezroczystości znaczników punktów pozwala zauważyć układ punktów przesłonięty przez chmurę przypadkowych danych.

Zmiana przezroczystości obszarów (za pomocą suwaka na dole wykresu) pozwala zobaczyć nakładające się części wykresu).



### 3.5. Linie odniesienia

Na wykresy można dużo łatwiej nanosić linie odniesienia, za pomocą wprowadzonej w wersji STATISTICA 10 karty *Linie odniesienia* w oknie ustawień wykresu.



### 3.6. Bezpośrednia edycja tekstów

Teksty umieszczone na wykresach można teraz edytować bez konieczności otwierania edytora. Edytor nadal jest dostępny i umożliwia wykonywanie bardziej złożonych operacji.

## 4. Środowisko pracy, ergonomia

### 4.1. Opis ogólny

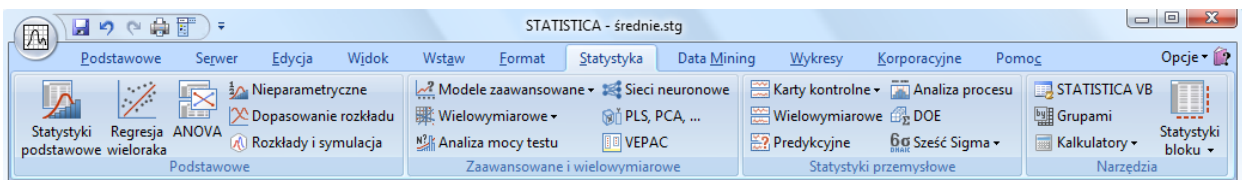
W programie STATISTICA 10 wprowadzono udoskonalenia do środowiska zwiększające komfort pracy użytkownika. Uwzględniają one ostatnie osiągnięcia ergonomii w dziedzinie ograniczenia

zmęczenia wzroku oraz udoskonalenia interakcji człowiek – komputer. *STATISTICA 10* ma lepsze i bardziej efektywne środowisko pracy, co zostało osiągnięte dzięki zmianie sposobu wyświetlania i nowym ikonom.

## 4.2. Wstążka

(Wszystkie produkty)

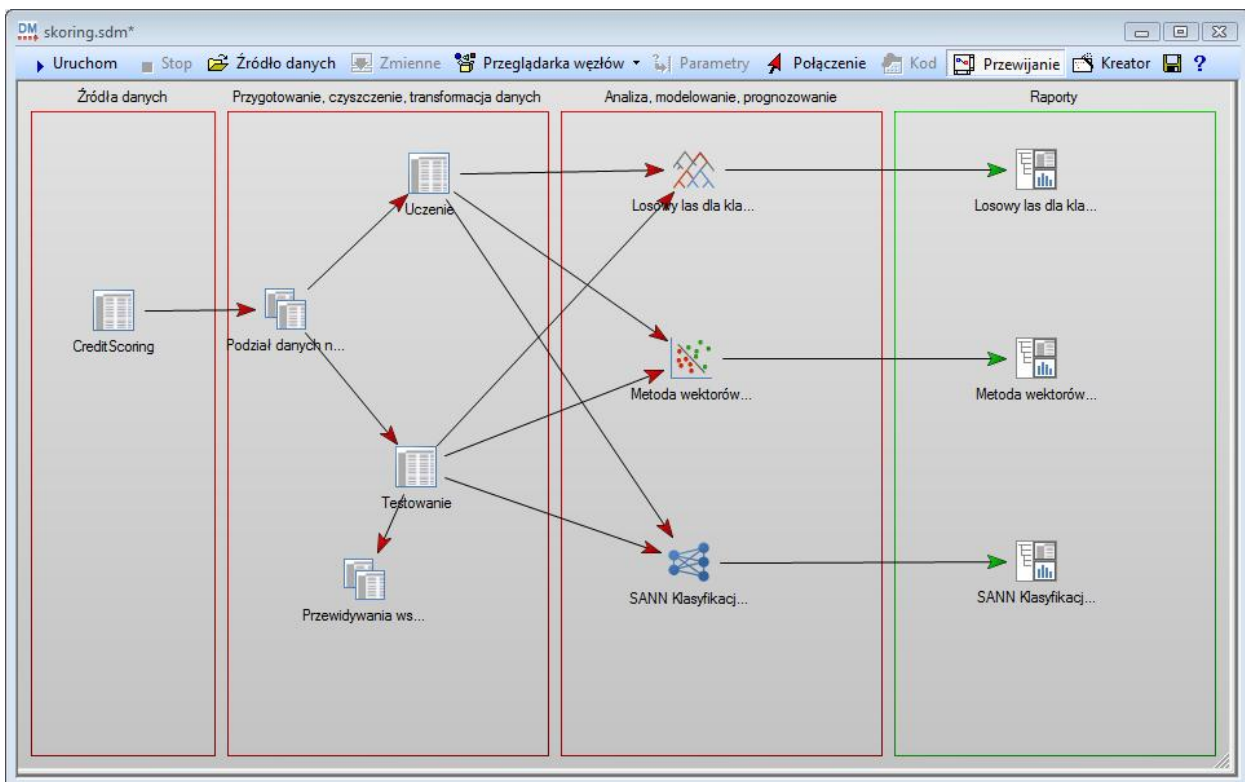
Wszystkie wstążki zostały odświeżone i zawierają zaprojektowane na nowo symbole (tradycyjne menu rozwijalne jest nadal dostępne dla zachowania zgodności). Makra *STATISTICA Visual Basic* można przypisywać do przycisków wstążki.



## 4.3. Przestrzeń robocza data mining

(*STATISTICA Data Miner*, *STATISTICA Text Miner*)

Ikony używane w przestrzeni roboczej data mining zostały powiększone i udoskonalono ich wygląd. Wprowadzono również inne nowe funkcje i udoskonalenia.



#### 4.4. Dodatki STATISTICA na wstążce

(Wszystkie produkty)

Programiści mogą dostosowywać wstążkę, korzystając z wywołań API (*Application Programming Interface*). Jest to w szczególności użyteczne przy tworzeniu dodatków STATISTICA.

### 5. Statystyki

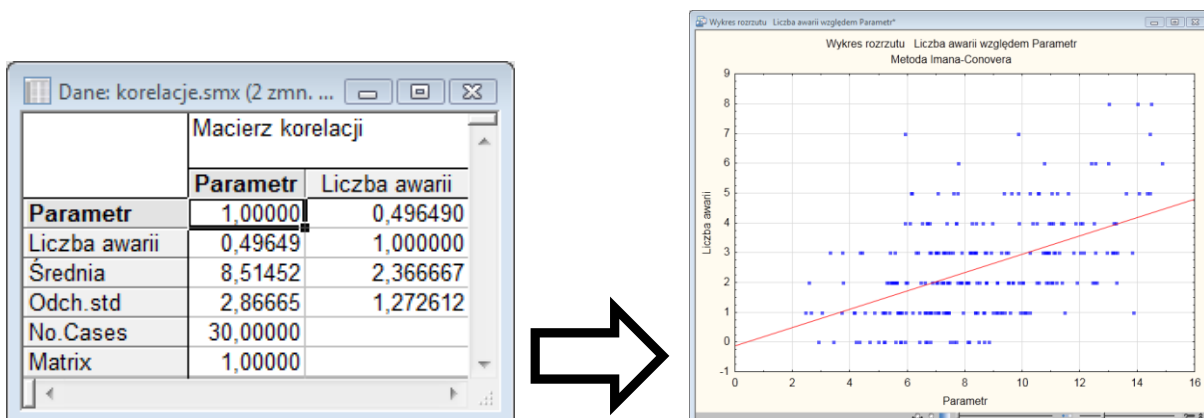
#### 5.1. Symulacja rozkładów i macierzy korelacji

(Wszystkie produkty zawierające *Pakiet Zaawansowany*)

Moduł STATISTICA *Rozkłady i symulacja*, wprowadzony w wersji 9, został ulepszony i rozszerzony. W STATISTICA 10 za pomocą nowego narzędzia *Układ symulacji* można łatwiej generować dane o wybranym rozkładzie.

Obecnie można znaleźć rozkład najlepiej pasujący do danych i na podstawie tego rozkładu i macierzy korelacji wygenerować podaną liczbę przypadków. Zamiast czekać na zgromadzenie odpowiedniej ilości danych, do zebranych danych dopasowujemy rozkład teoretyczny, następnie przeprowadzamy symulację i na jej podstawie i wyciągamy wnioski. Przy symulacji możemy uwzględnić korelacje zmiennych. Narzędzie to jest niezwykle użyteczne w analizach „co – jeśli” i jest coraz częściej stosowane w różnych dziedzinach.

Dla przykładu rozważmy wytwarzanie skomplikowanej maszyny. Wiedza o tej maszynie i jej częściach może zostać wykorzystana do wygenerowania danych. Dane te z kolei mogą być użyte w badaniach niezawodności. Poniżej widzimy macierz korelacji pewnego parametru pracy maszyny i liczby awarii. Macierz ta została wyznaczona na podstawie danych z przeszłości. Ponadto wyznaczono również średnie i odchylenia standardowe, a także dopasowano rozkłady teoretyczne. Używając *Układu symulacji*, generujemy wartości każdej zmiennej zgodnie z ich rozkładem i macierzą korelacji. Użytkownik może wybierać rodzaj rozkładu i określać jego parametry albo skorzystać z wartości dopasowanych do danych. Wygenerowane dane przedstawia wykres rozrzutu. Możemy je wykorzystać do rozmaitych analiz dotyczących niezawodności maszyny, optymalizacji warunków pracy itd.



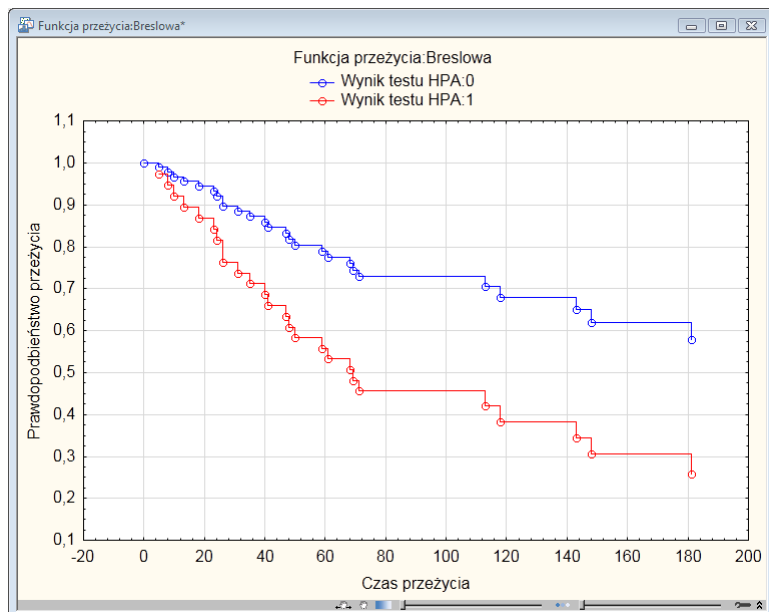
Wielowymiarowa symulacja może być wykorzystana również w podejściu [Quality by Design](#), stosowanym w przemyśle farmaceutycznym i zalecanym przez EMA i FDA.

## 5.2. Modele proporcjonalnego hazardu Coxa

(Wszystkie produkty zawierające *Pakiet Zaawansowany*)

W wersji 10 wprowadzono rozbudowaną i wysoce skalowalną procedurę *Modele proporcjonalnego hazardu Coxa*, będącą potężnym narzędziem do modelowania danych o czasie trwania. Zastosowania tego modułu to m.in.:

- Analiza przeżycia (badania medyczne),
- Przewidywanie odejść klientów (*churn*),
- Modelowanie czasu bezawaryjnej pracy urządzeń (niezawodność).



Moduł *Modele proporcjonalnego hazardu Coxa* umożliwia pracę z danymi uciętymi, uwzględnianie jakościowych zmiennych objaśniających (predyktorów), uwzględnienie interakcji i zagnieżdżenia. Model można tworzyć metodami najlepszego podzbioru i krokową, a następnie stosować go za pomocą modułu *STATISTICA Szybkie wdrażanie*.

## 5.3. Dalsze udoskonalenia wydajności i skalowalności w obliczaniu statystyk opisowych

(Wszystkie produkty)

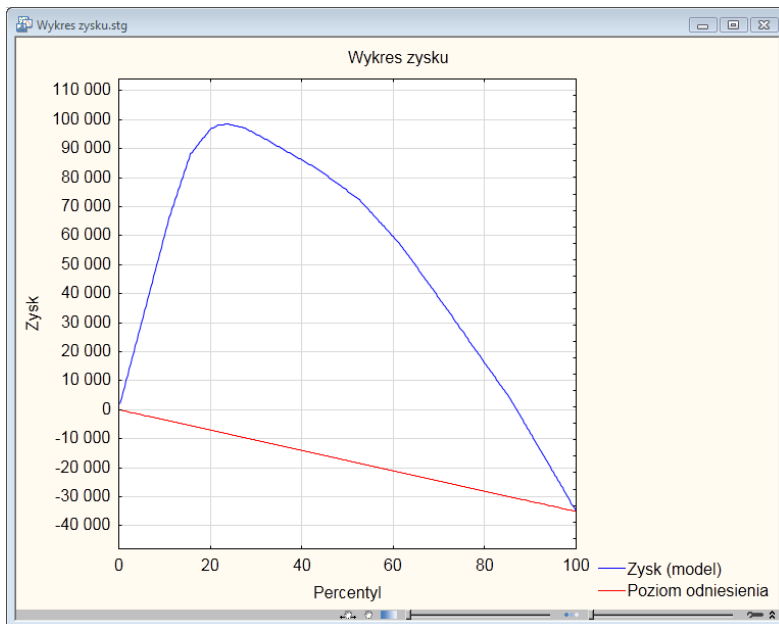
Dokonano wielu udoskonalień w obliczaniu statystyk opisowych, dzięki czemu w wielu przypadkach uzyskano znaczące przyspieszenie obliczeń dla bardzo dużych zbiorów danych. Przykładowo wielowątkowe wyznaczanie przekrojów (w tym percentyli) zostało udoskonalone, aby uzyskać nadzwyczajną szybkość obliczeń dla bardzo dużych zbiorów danych.

## 6. Data Mining

### 6.1. Wykres zysku

(STATISTICA Data Miner)

W module *Szybkie wdrażanie modeli predykcyjnych* dostępny jest wykres zysku. Pokazuje on np. jaki będzie zysk z kampanii marketingowej, jeśli dobierzemy do niej osoby wskazane przez model z wybranym prawdopodobieństwem. Wykres ma wiele zastosowań jako jedno z narzędzi oceny jakości modelu i doboru punktu odcięcia.



### 6.2. Krzywa ROC

(STATISTICA Data Miner)

Za pomocą modułu *Szybkie wdrażanie modeli predykcyjnych* można teraz tworzyć krzywą ROC. Jest ona użytecznym narzędziem oceny trafności modelu, pokazującym, jak często model trafnie wskazał obiekty należące do klasy, i jednocześnie, jak często takie wskazanie było błędne. Krzywa ROC jest wykorzystywana w wielu dziedzinach, takich jak: medycyna, psychologia i sterowanie jakością.

### 6.3. Dalsze udoskonalenia wydajności i skalowalności metod predykcyjnych

(STATISTICA Data Miner)

Uzyskano dalszą, znaczącą poprawę wydajności rozmaitych algorytmów modelowania predykcyjnego dla bardzo dużych zbiorów danych. Przykładowo tworzenie wszystkich uogólnionych modeli liniowych (np. regresji logistycznej) odbywa się z wykorzystaniem wielu rdzeni lub procesorów i obsługą bardzo dużych zbiorów danych. Podobne udoskonalenie skalowalności i znaczącą poprawę wydajności osiągnięto dla algorytmów C&RT i CHAID.

### 6.4. Text Mining

(STATISTICA Text Miner)

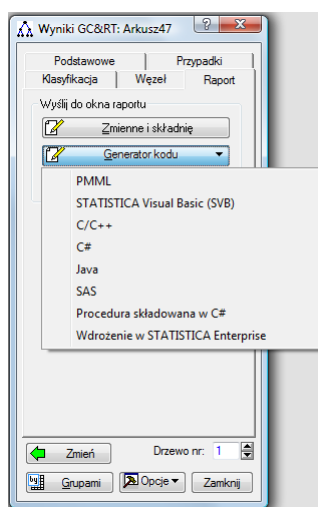
W odpowiedzi na ostatnie trendy w statystycznej analizie dokumentów tekstowych, gdzie badane i modelowane są ogromne zbiory danych, znacząco przebudowano i jeszcze bardziej zoptymalizowano główny silnik obliczeniowy *STATISTICA Text Miner*, aby zwiększyć skalowalność

i wydajność. Wewnętrzna procedura obsługi bazy danych została przeprojektowana i może wydajnie obsługiwać ekstremalnie duże zbiory danych z intensywnym wykorzystaniem wielowątkowości.

## 6.5. Wdrożenie w Java i C#

(*STATISTICA Data Miner* – wdrożenie w bazie danych)

*STATISTICA 10* dysponuje dwoma nowymi narzędziami do wdrożenia modeli w postaci kodu Java i C#. Modele mogą być zapisywane w kodzie C#, który można bezpośrednio wbudować w funkcje MS SQL Server, stosowane później jako procedury składowane wyznaczające wartości wynikające z modelu bezpośrednio wewnątrz bazy danych. W podobny sposób kod Java może być wykorzystywany w bazie Oracle. Korzystanie z tych narzędzi wymaga specjalnej licencji. Główną zaletą stosowania modeli bezpośrednio w bazie danych jest zwiększenie wydajności: stosowanie modelu wewnątrz bazy może być wykonywane nawet o rząd wielkości szybciej niż przetwarzanie zewnętrzne.



## 6.6. STATISTICA Zestaw Skoringowy

(*STATISTICA Zestaw Skoringowy*)

Narzędzia do tworzenia tabel skoringowych są teraz w pełni zintegrowane ze *STATISTICA* i zostały dodatkowo rozbudowane.

*STATISTICA Zestaw Skoringowy* stanowi dedykowane rozwiązanie do tworzenia, oceny i monitorowania tablic skoringowych. Program umożliwia wybór predyktorów, dyskretyzację zmiennych, budowę modelu dla tablicy skoringowej (model logistyczny), dobór punktu odcięcia, badanie wpływu wniosków odrzuconych, badanie stabilności populacji i cech oraz analizę Vintage.

W nowej wersji (3) dodano dwa moduły: modele skoringowe typu SURVIVAL oraz analizę Vintage. Dodatkowo wprowadzono szereg zmian i udoskonalień, z których najważniejsze to:

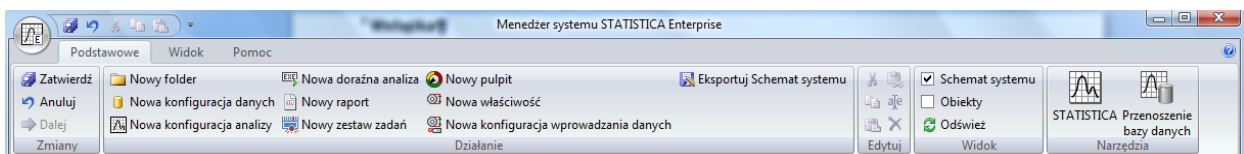
- 1) Wybór predyktorów
  - Dodatkowe algorytmy rotacji czynników przy wyborze reprezentantów
  - Możliwość uwzględnienia wartości nietypowej podczas budowy rankingu predyktorów
- 2) Budowa tablicy skoringowej
  - Możliwość korekty punktacji w gotowej tablicy skoringowej, także wczytanej za pomocą pliku XML
  - Zapis tablicy skoringowej do pliku Excela
  - Raport z przebiegu doboru zmiennych do modelu w przypadku regresji krokowej
  - Miary dopasowania modelu - R2 McFaddena, Coxa-Snella oraz Cragga-Uhlera

- 3) Ocena modeli
  - Możliwość oceny modeli równocześnie na zbiorze uczącym i testowym
  - Możliwość obliczenia wskaźnika Lift dla obydwu modelowanych klas
  - Możliwość uwzględniania modelu SURVIVAL
  - Wartości p dla testów KS oraz Hosmera-Lemeshowa
- 4) Obliczanie skoringu
  - Obliczanie skoringu dla modeli typu SURVIVAL
  - Symulacja przebiegu funkcji przeżycia dla różnych wartości parametrów
- 5) Zarządzanie punktem odcięcia
  - Możliwość wyboru punktu odcięcia na podstawie jednostkowych bądź średnich kosztów (np. wyrażonych kwotowo) błędniej klasyfikacji oraz zysków poprawnej klasyfikacji
- 6) Stabilność populacji
  - Standardowe raporty stabilności populacji i cech zostały uzupełnione o zestaw wykresów pozwalających na łatwiejszą identyfikację przyczyn odejścia od stabilności.

## 7. STATISTICA Enterprise

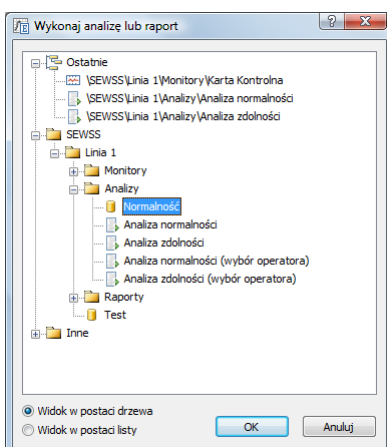
### 7.1. Wstążka

Praca w *Menadżerze systemu STATISTICA Enterprise* jest prostsza i bardziej efektywna dzięki wprowadzeniu wstążki.



### 7.2. Konfiguracje danych

Konfiguracje danych można uruchomić z poziomu *STATISTICA*, dzięki czemu użytkownik może przeglądać dane bez konieczności uruchamiania *Menadżera systemu STATISTICA Enterprise*.



### 7.3. Przenoszenie bazy danych

Narzędzie do przenoszenia systemowej bazy danych zostało zaktualizowane i jest dostępne w *Menadżerze systemu STATISTICA Enterprise*. Za jego pomocą administrator może przenosić konfigurację z jednej systemowej bazy danych do drugiej.

#### 7.4. Wdrażanie makr w STATISTICA Enterprise

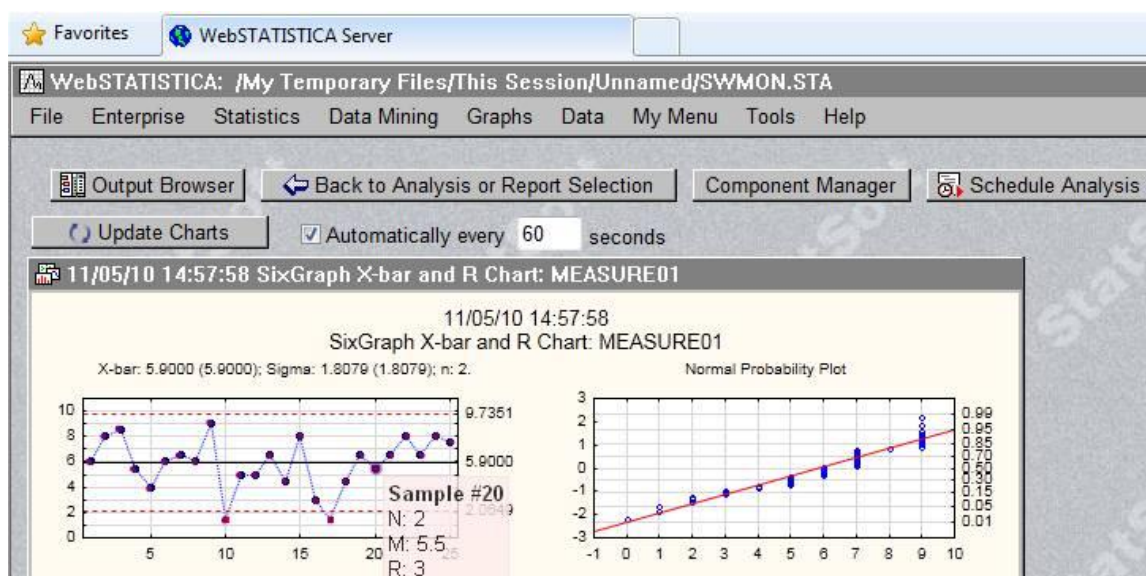
STATISTICA 10 ułatwia korzystanie z makr w STATISTICA Enterprise. Jest to prostsze niż ręczne tworzenie konfiguracji analiz SVB, a działa nie tylko dla makr, ale również skryptów R. Aby wdrożyć makro, należy po jego utworzeniu w STATISTICA nacisnąć przycisk *Wdróż makro* na karcie *Korporacyjne wstążki*.

#### 7.5. Nazwy konfiguracji

*Menadżer systemu* daje teraz większą swobodę w nazywaniu obiektów. Nazwy muszą być unikalne tylko w obrębie jednego folderu systemu.

#### 7.6. STATISTICA Enterprise Server - automatyczna aktualizacja kart kontrolnych

Karty kontrolne (tworzone jako konfiguracje analizy typu SPC) z włączoną automatyczną aktualizacją są teraz automatycznie odświeżane, również gdy są otwierane w przeglądarce internetowej. Użytkownik, korzystając z przeglądarki, może zmieniać odstęp między aktualizacjami lub ręcznie je uruchomić. Wykorzystywane są najnowsze technologie internetowe, pozwalające odświeżyć wykres bez ładowania całej strony (tzn. bez efektu migania strony).



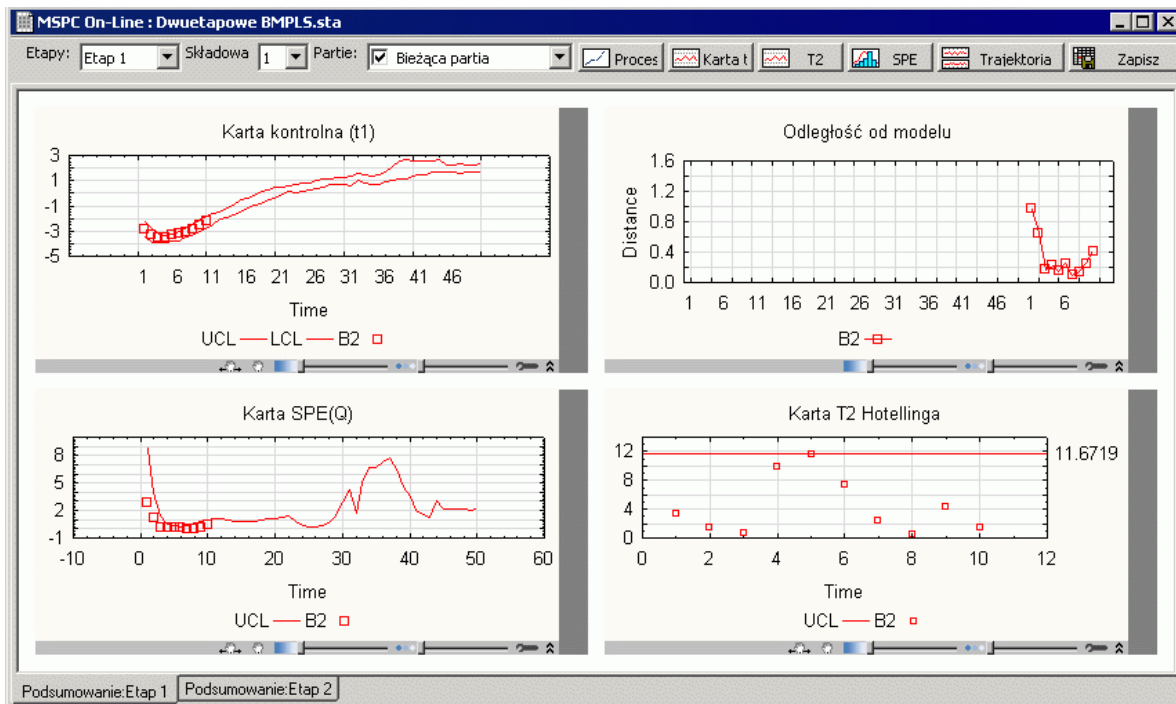
#### 7.7. STATISTICA Enterprise Server - eksploracja karty kontrolnej

Karty kontrolne uruchamiane w środowisku przeglądarkowym mogą być interaktywnie eksplorowane. Możemy przypisywać punktom na karcie przyczyny, działania i komentarze, a także włączać i wyłączać ich uwzględnianie w obliczeniach i na wykresach. Wykorzystywane są w nich najnowsze technologie internetowe, pozwalające odświeżyć wykres bez ładowania całej strony (tzn. bez efektu migania strony).

## 8. STATISTICA MSPC Online

(STATISTICA MSPC Online)

W STATISTICA 10 możemy łatwiej stosować analizy wielowymiarowe (PCA, PLS) dzięki nowemu narzędziu STATISTICA MSPC Online. Umożliwia ono automatyczną aktualizację, monitorowanie i interakcyjne drążenie danych poprzez przechodzenie od wartości składowych do wykresów wkładu i kart jednowymiarowych.



## 9. STATISTICA Web Data Entry

(STATISTICA Web Data Entry)

STATISTICA Web Data Entry umożliwia użytkownikom określenie okien wprowadzania danych wyświetlanych w przeglądarce internetowej. Wpisane dane gromadzone są w systemowej bazie STATISTICA Enterprise.

STATISTICA 10 Web Data Entry zawiera wiele udoskonaleń, takich jak:

- łatwe do skonfigurowania pola wymagane
- Udoskonalona nawigacja
- Układanie pól w grupy dla ułatwienia wprowadzania danych
- Znajdowanie wprowadzonych wcześniej próbek według etykiet
- Udoskonalone pobieranie danych na potrzeby analiz

## 10. Dedykowane rozwiązania skoringowe

### 10.1. STATISTICA Zestaw Skoringowy

(STATISTICA Zestaw Skoringowy)

Dostępne już we wcześniejszych wersjach narzędzia do tworzenia modeli i kart skoringowych są teraz w pełni zintegrowane ze *STATISTICA* i zostały dodatkowo rozbudowane. Więcej informacji w punkcie 6.6.

## **10.2. STATISTICA Live Score**

### 11. (*STATISTICA Live Score*)

Z wersją 10 wprowadzono nowe i udoskonalone wydanie *STATISTICA Live Score*. *STATISTICA Live Score* jest to rozwiązanie wykorzystujące *STATISTICA Enterprise Server* i narzędzia statystyki i data mining *STATISTICA*. W *STATISTICA Data Miner* przygotowywane są dane (agregowane, czyszczone itd.), tworzone i oceniane modele. Po sprawdzeniu i zatwierdzeniu modeli są one stosowane za pomocą serwera *STATISTICA Live Score*. Oblicza on i udostępnia wartości skoringu wynikające z modelu, wykorzystując zaawansowaną, wielowątkową architekturę. Przykłady zastosowań:

- Wyznaczanie tablic skoringowych na potrzeby obsługi klienta (np. biura obsługi klienta i rozwiązania internetowe).
- Wyznaczanie segmentu, do którego należy dana osoba, wspomaganie sprzedaży wiązanej i rozszerzonej, zapobieganie odejściom klienta itp.
- Alarmowanie na potrzeby wykrywania i zapobiegania nadużyciom.

## **12. Pomoc i dokumentacja**

(Wszystkie produkty)

### **Przykłady modelu obiektowego STATISTICA**

Do *Pomocy* dodano setki przykładów stosowania modelu obiektowego *STATISTICA Visual Basic*.