



## **STATISTICA ENTERPRISE JAKO PLATFORMA ANALITYCZNA DLA CAŁEJ ORGANIZACJI**

*Tomasz Demski, StatSoft Polska Sp. z o.o.*

### **Co to jest STATISTICA Enterprise?**

*STATISTICA Enterprise* to specjalna wersja *STATISTICA* przeznaczona do wykorzystania przez zespoły ludzi, pracujących wspólnie i potrzebujących wygodnego, wydajnego i bezpiecznego środowiska do analizy danych i raportowania. Ponieważ *STATISTICA Enterprise* łączy w sobie narzędzia do przeprowadzenia całego procesu przekształcania surowych danych w użyteczną wiedzę, nazywamy go zintegrowanym systemem analizy danych.

W skład systemu wchodzi:

- ◆ centralna baza danych,
- ◆ narzędzia analityczne *STATISTICA* (statystyka, wizualizacja i *data mining*),
- ◆ generator raportów,
- ◆ system bezpiecznego przechowywania dokumentów zgodny z wymaganiami FDA CFR Część 11 (ta część jest opcjonalna).

Poglądowy schemat działania systemu i główne role użytkowników przedstawia rys. 1.

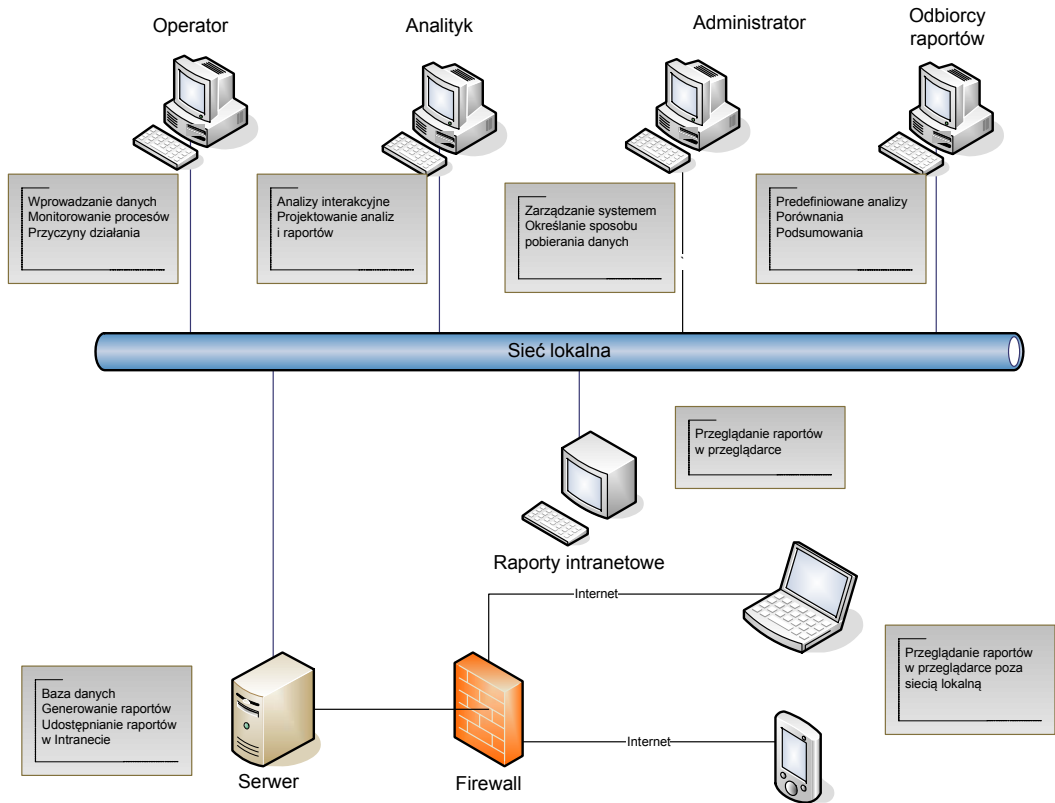
Rdzeniem zintegrowanego systemu jest centralna baza danych. Możemy w niej przechowywać:

- ◆ wyniki pomiarów właściwości procesu i towarzyszące im dodatkowe informacje (etykiety), takie jak np. identyfikator osoby wprowadzającej konkretne obserwacje,
- ◆ specyfikacje sposobu pobierania danych z zewnętrznych źródeł danych (baz danych, plików, sterowników automatyki przemysłowej i innych),
- ◆ scenariusze wprowadzania danych (mierzone właściwości, dopuszczalne wartości, urządzenia pomiarowe itd.),
- ◆ scenariusze analizy danych i szablony raportów,
- ◆ uprawnienia dostępu użytkowników do poszczególnych funkcji i danych systemu.

Warto zwrócić uwagę na kluczową zaletę przechowywania powyżej wymienionych danych w centralnej bazie danych. Dzięki temu mamy zapewnioną dostępność i zgodność danych, a potrzebne ustawienia i zmiany wystarczy wykonać raz. Jeśli przykładowo zmieniamy



szablon raportu, np. dodając do niego nowy wykres, to wystarczy wykonać to raz i zapisać nowe ustawienia w bazie danych. Zmiana ta jest potem automatycznie przenoszona na wszystkie stanowiska, nie trzeba obchodzić każdego komputera i zmieniać na nim ustawień. Ponadto, dane wprowadzone na jednym stanowisku są dostępne dla użytkowników na innych stanowiskach praktycznie od razu.



Rys. 1. Schemat *STATISTICA Enterprise*.

W skład *STATISTICA Enterprise* wchodzi narzędzia analizy danych z rodziny *STATISTICA*. Zakres metod statystycznych, wizualizacji i *data mining* zawartych w systemie zależy od potrzeb danego użytkownika. Z narzędzi analitycznych korzystamy na kilka sposobów (przy czym wiele osób nie tylko nie musi umieć obsługiwać programu *STATISTICA*, ale nawet wiedzieć, że jest źródłem uzyskiwanych informacji), są to:

- ♦ interaktywne wykonywanie analiz (tak jak w desktopowej wersji *STATISTICA*), projektowanie szablonów analiz i raportów dla innych użytkowników,
- ♦ korzystanie z przygotowanych scenariuszy analizy w środowisku *STATISTICA*,
- ♦ przeglądanie raportów (np. w przeglądarce internetowej lub PDF),

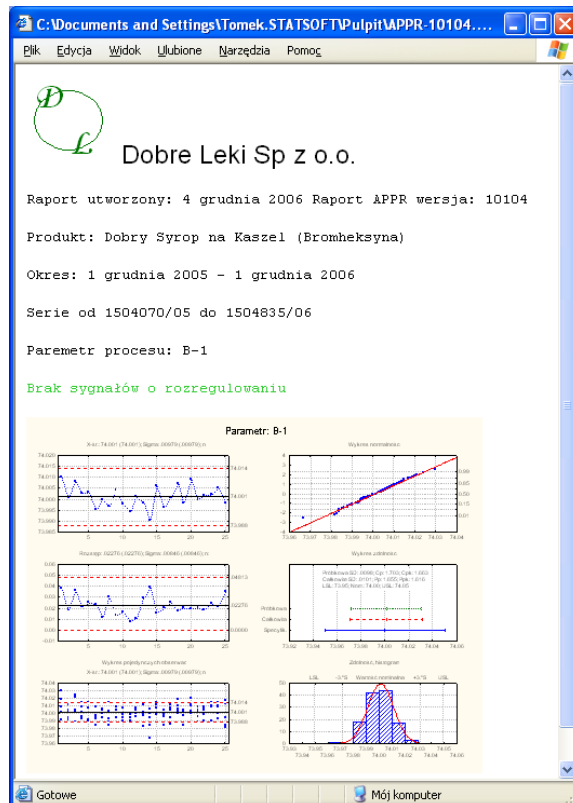


- ◆ odbiór automatycznie generowanych przez system powiadomień (alarmów) poprzez e-mail, sms itp.

Z praktycznego punktu widzenia warto zwrócić uwagę na aspekt łatwości korzystania z wyników analiz w *STATISTICA Enterprise*. Otóż w praktyce najczęściej większość osób nie wykonuje analiz, a jest odbiorcą ich wyników. Takim osobom dużo łatwiej jest korzystać z szablonów analizy przygotowanych przez specjalistów od analizy danych i wdrożonych w *STATISTICA Enterprise*, niż własnoręcznie określać i uruchamiać analizę w zwykłym (tj. niewbudowanym w system) programie do analizy danych.

Generator raportów służy do uruchamiania całej analizy w zadanych odstępach czasu i przekształcania jej wyników na powszechnie zrozumiały format (np. PDF) – tak, aby mógł z nich korzystać praktycznie każdy. Układ i formatowanie raportu określane jest podczas jego tworzenia przez użytkownika. Oprócz wyników analizy (tabel, wykresów i informacji tekstowych) raport może być wzbogacony:

- ◆ dodatkowymi danymi: datą i czasem utworzenia, nazwą użytkownika, który uruchomił raport, wersją raportu itp.,
- ◆ tekstami (np. tytułem) i obiektami graficznymi (np. logo, schemat produkowanego elementu, którego dotyczy analiza).



Rys. 2. Przykładowy raport *STATISTICA Enterprise*.



Generator raportów jest w stanie tworzyć raport spełniający nawet bardzo wygórowane wymagania dotyczące jego zawartości, ale również bezpieczeństwa danych i monitorowania wszelkich zmian. Przykładowy raport przedstawia rys. 2.

Generator raportów może w praktyce okazać się niezwykle pożyteczny. Standaryzacja i automatyzacja tworzenia zestawień pozwala spożytkować czas analityków na twórcze wykorzystanie wiedzy o procesie i jego doskonalenie, zamiast na wykonywanie żmudnych i powtarzalnych czynności. Zastosowanie takiego systemu oszczędza tysiące godzin pracy rocznie (przykładowo przygotowanie zestawienia podsumowującego w jednej z firm farmaceutycznych przed wdrożeniem systemu bazującego na *STATISTICA Enterprise* trwało 40 godzin, a po wdrożeniu 5 min!). Oprócz bezpośredniej oszczędności czasu i środków, zastosowanie systemu zmniejsza ryzyko popełnienia błędu i daje bardziej dokładny wgląd w proces, poprzez dostępność bardziej złożonych metod analitycznych.

Samo przechowywanie wszystkich danych w centralnej bazie danych oraz system uprawnień użytkowników zapewnia bezpieczeństwo danych. Zwłaszcza że *STATISTICA Enterprise* dysponuje funkcją zapisu działań użytkowników systemu. Jednak w pewnych zastosowaniach jest to za mało – wymagane jest:

- ◆ bezpieczne przechowywanie dokumentów tworzonych przez system,
- ◆ funkcja zatwierdzania dokumentów,
- ◆ kontrola i zapis zmian (innymi słowy przechowywanie wszystkich wersji dokumentów, szablonów analiz i raportów).

W takim wypadku potrzebna jest specjalna wersja *STATISTICA Enterprise* zintegrowana z *STATISTICA Document Management System* (opis tego systemu znajduje się na stronie <http://www.statsoft.pl/dms.html>). Ta wersja *STATISTICA Enterprise* jest zgodna z wymaganiami FDA CFR Część 11 i nosi nazwę *Compliance Edition*.

## Zastosowania *STATISTICA Enterprise*

Zintegrowany system analizy danych ma zastosowanie wszędzie tam, gdzie z wiedzy wydobytej z danych będzie korzystało wiele osób współpracujących ze sobą, lecz mających odmienne kompetencje, zadania i uprawnienia.

Wdrożenie systemu jest także pożyteczne, jeżeli powinniśmy w regularnych odstępach czasu tworzyć rutynowe raporty i podsumowania zawierające statystyki i wykresy (o jednym z takich zastosowań wspomnieliśmy na str. 78).

Najłatwiej dostrzec potrzebę zintegrowanego systemu analizy danych w zastosowaniach związanych z jakością produkcji lub usług. W tej dziedzinie normy (np. ISO 9000, ISO/TS 16949) oraz strategię zarządzania (Six Sigma, TQM) wymagają stosowania metod statystycznych i korzystania z nich nieomal przez wszystkich pracowników. Analizy powinny być wykonywane na bieżąco w toku procesu. Takie wymagania dokładnie odpowiadają możliwościom *STATISTICA Enterprise*.



Inna dziedzina zastosowań to wspieranie analiz związanych z CRM, dotyczących takich zadań jak: segmentacja, migracja klientów (*churn*), sprzedaż krzyżowa (*cross-selling*) i rozszerzona (*up-selling*), optymalizacja kampanii marketingowych itp. Analizy związane z CRM wymagają zazwyczaj pobierania danych z wielu źródeł danych, łączenia tych danych i skomplikowanych przekształceń. Wszystkie te operacje możemy zapisać w bazie *STATISTICA Enterprise* jako konfigurację danych, co znakomicie ułatwia właściwą analizę, a potem stosowanie jej wyników.

Z analiz na potrzeby CRM korzysta wiele osób, z których większość korzysta wyłącznie z wyników analiz, np. pracownik telecentrum otrzymuje listę klientów, do których powinien zadzwonić, bo właśnie dla nich ryzyko odejścia do konkurencji jest największe. Potrzeby tego typu użytkowników można zaspokoić, korzystając funkcji systemu automatycznego uruchamiania analizy i zapisywania jej wyników w wybranym formacie.

W zastosowaniach związanych z CRM bardzo użyteczna jest również możliwość łatwego stosowania modeli dla nowych danych.

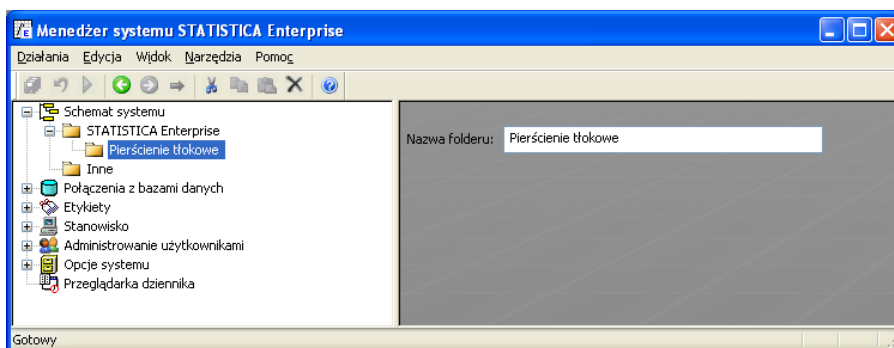
Inna sfera zastosowań *STATISTICA Enterprise* to prognozowanie, zwłaszcza gdy musimy przewidywać wiele zmiennych (np. przewidywanie sprzedaży różnych produktów w wielu regionach). Bardzo przydatne jest tu automatyczne stosowanie modelu i monitorowanie jakości przewidywań. Podobnie jak w CRM, również do praktycznego stosowania prognoz wymagane jest dostarczanie przewidywań osobom podejmującym decyzje, które nie muszą znać się na samych technikach prognozowania i umieć stosować ich w programie do analizy danych. *STATISTICA Enterprise* umożliwia łatwe spełnienie tego wymogu.

Są zastosowania (np. dotyczące projektowania i produkowania leków), w których musimy zapewnić maksymalne bezpieczeństwo danych, uniemożliwić fałszowanie wyników oraz dokumentować wszystkie działania użytkownika, w szczególności zmian dokonywanych w specyfikacjach analiz i raportów. Te wymogi spełnia *STATISTICA Enterprise Compliance Edition*, a spełnienie ich w inny sposób jest trudne i w praktyce wymaga zbudowania analogicznego systemu we własnym zakresie.

## Przykład pracy w *STATISTICA Enterprise*

Zobaczymy teraz, jak wygląda praca w systemie *STATISTICA Enterprise*. Naszym zadaniem jest utworzenie karty kontrolnej dla średnicy pierścieni tłokowych. Wyniki pomiarów zapisane są w bazie danych, a więc będziemy musieli zacząć od określenia sposobu pobierania danych. Karty kontrolnej użyjemy do zdiagnozowania procesu, a jeśli wszystko będzie w porządku, to zapiszemy ją jako narzędzie do monitorowania procesu przez operatorów.

Zaczynamy od zalogowania się w systemie. Możemy podać nazwę użytkownika i hasło zdefiniowane w *STATISTICA Enterprise* lub zalogować się automatycznie, używając swojej nazwy użytkownika z systemu Windows.




Rys. 3. Menedżer systemu *STATISTICA Enterprise*.

Po zalogowaniu na ekranie pojawi się *Menedżer systemu* prezentujący obiekty systemu, z których możemy korzystać (rys. 3). Okno tej aplikacji podzielone jest na dwa panele: w lewym widzimy uporządkowany hierarchicznie schemat systemu, a w prawym mamy ustawienia dla aktualnie podświetlonego obiektu. Panel schematu systemu obsługujemy analogicznie jak *Eksplorator Windows*.

Pracę rozpoczynamy od utworzenia nowej **konfiguracji danych**. Konfiguracja danych to szczegółowy przepis na pobieranie danych z zewnętrznego źródła danych. W skład takiego przepisu wchodzi informacja: skąd pobierać dane, które wielkości odczytywać, jakie kryteria mają one spełniać (np. parametry procesu z ostatniego okresu) oraz jaka jest interpretacja wielkości (np. wartości pola *Pomiar* to właściwość kreślona na karcie kontrolnej).

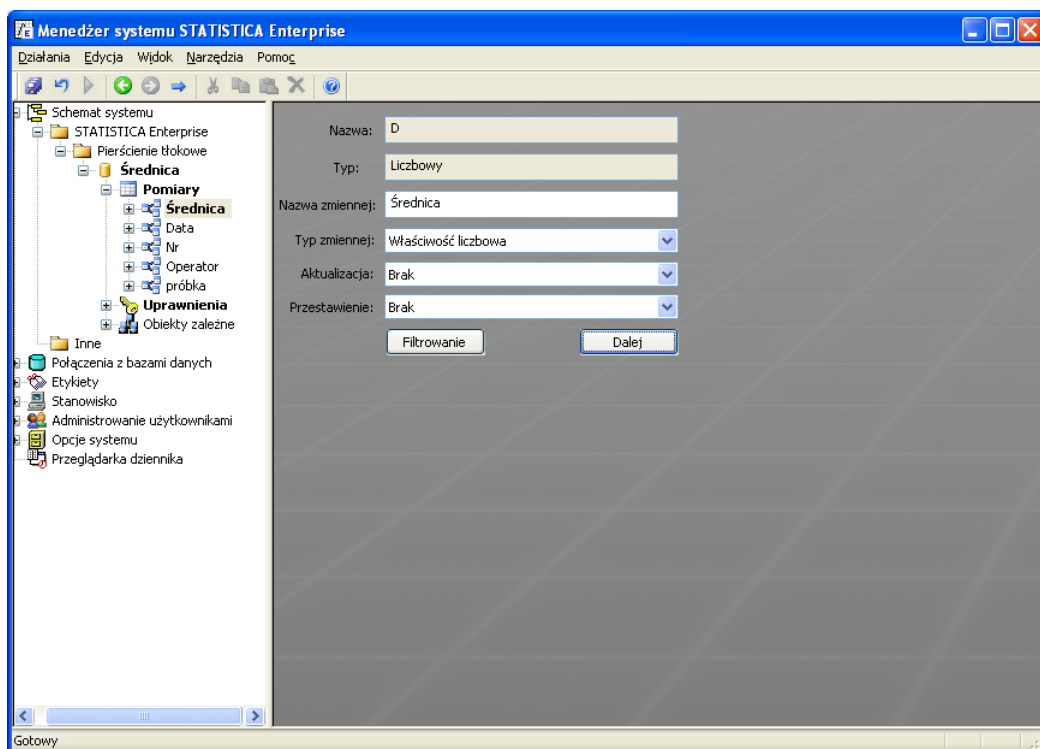
Nową konfigurację danych umieścimy w folderze *Pierścienie tłokowe*. Podświetlamy go w schemacie systemu (tak jak na rys. 3) i z menu *Działania* wybieramy polecenie *Nowa konfiguracja danych*. Następnym krokiem jest nazwanie tworzonego obiektu: ponieważ będziemy pobierać dane o średnicy pierścieni tłokowych, konfigurację nazwiemy *Średnica*. Następnie z listy *Połączenia* wybieramy połączenie do bazy danych *Pierścienie tłokowe* (połączenie określa, z jaką bazą danych się łączymy i w jaki sposób, m.in. z jakiego sterownika będziemy korzystać).

Po wybraniu połączenia klikamy przycisk *Dalej*. Program poprosi nas o określenie zapytania do bazy danych. Zapytanie można wpisać ręcznie lub skorzystać z *Kreatora SQL*, co zazwyczaj jest znacznie wygodniejsze. Po kliknięciu przycisku *Kreator SQL* na ekranie otworzy się okno *STATISTICA Query* (rys. 4), w którym w trybie graficznym możemy określić zapytanie do bazy danych. W naszym przypadku wybieramy wszystkie pola z jedynej dostępnej dla nas tabeli bazy danych i klikamy przycisk , aby zapisać zapytanie i powrócić do *Menedżera systemu*.





Teraz powinniśmy określić, jak w *STATISTICA* traktowane będą dane odczytywane z bazy danych. Możemy też nadać zmiennym nazwy różniące się od nazw pól w bazie danych. Zaczniemy od kolumny *D*: klikamy odpowiadający mu wiersz w tabeli *Kolumny OLE DB* i naciskamy przycisk *Edytuj*. Przejdziemy do panelu właściwości dla zmiennej odczytywanej z pola *D* w bazie danych. W panelu tym w polu *Nazwa zmiennej* wpisujemy *Średnica* (pod tą nazwą wielkość ta będzie pojawiała się w *STATISTICA*), a z listy *Typ zmiennej* wybieramy *Właściwość liczbowa*, aby program wiedział, że dla tej zmiennej ma tworzyć karty kontrolne. Ustawienia te widzimy na rys. 6.



Rys. 6. Właściwości zmiennej dla kolumny *D*.

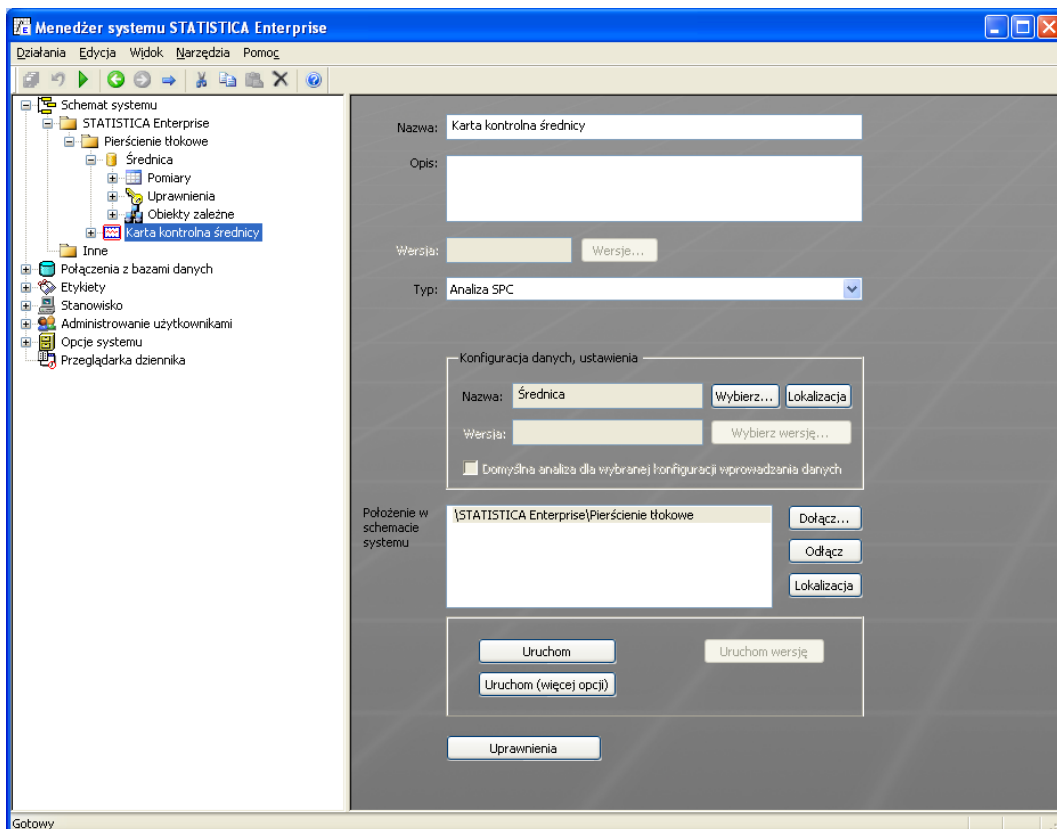
W analogiczny sposób określamy sposób interpretacji dla pozostałych kolumn:

- ♦ dla kolumny *próbka* jako *Typ zmiennej* wybieramy *Identyfikator próbki*,
- ♦ dla kolumny *Nr* jako *Typ zmiennej* wybieramy *Identyfikator jednostki*,
- ♦ dla kolumny *Data* jako *Typ zmiennej* wybieramy *Oznaczenia czasu* i z listy *Aktualizacja* wybieramy *Pierwsza przy aktualizacji*, aby program na podstawie wartości tej zmiennej sprawdzał, czy pojawiły się nowe próbki i czy należy zaktualizować dane i analizy,
- ♦ dla kolumny *Operator* jako *Typ zmiennej* wybieramy *Etykieta* (zmienna ta będzie traktowana jako kryterium filtrowania).



Po wykonaniu powyższych czynności do zakończenia procesu definiowania konfiguracji danych wystarczy określić, kto może z niej korzystać – robimy to po kliknięciu węzła *Uprawnienia* dla konfiguracji *Średnica*.

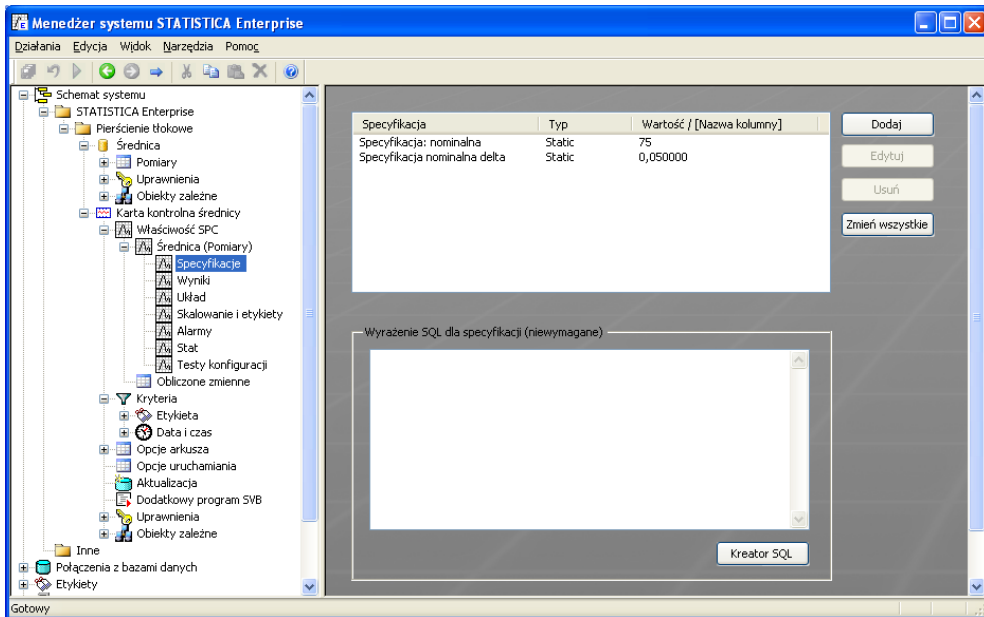
Mamy gotowe źródło danych do naszej analizy i możemy rozpocząć tworzenie samej analizy. Podświetlamy konfigurację *Średnica* i z menu *Działania* wybieramy polecenie *Nowa konfiguracja analizy dla tych danych*. Program utworzy nową konfigurację danych i wyświetli panel z jej ogólnymi właściwościami. Zmieniamy w nim tylko nazwę konfiguracji na *Karta kontrolna średnicy* (rys. 7).



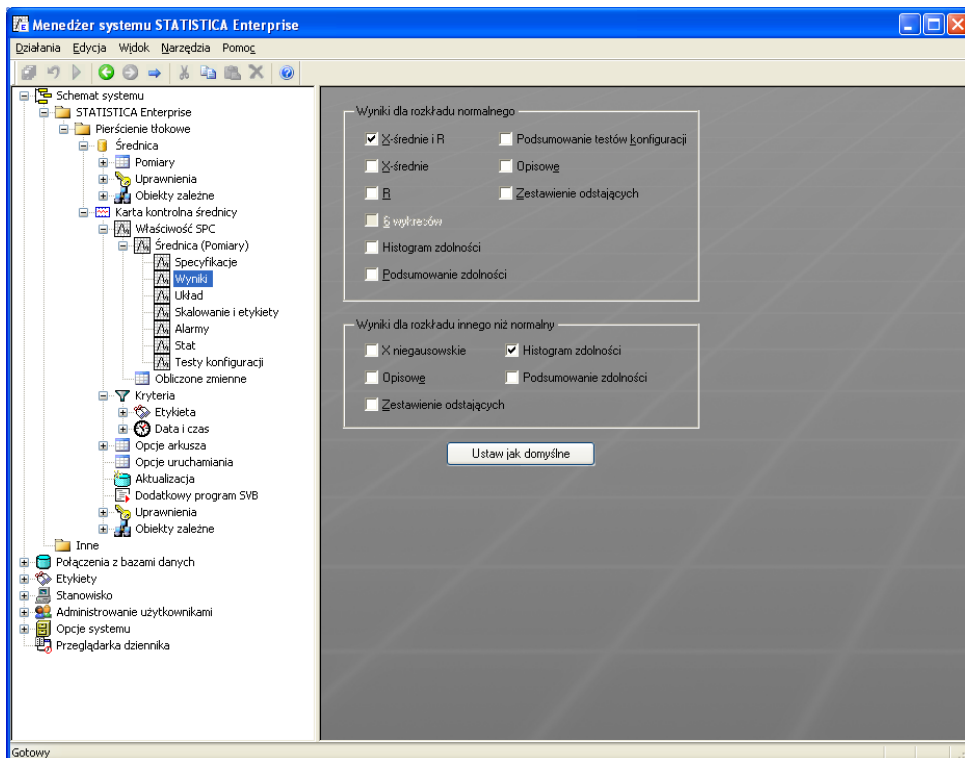
Rys. 7. Ogólne właściwości konfiguracji analizy.

Parametry analizy ustawiamy dla każdej zmiennej określonej jako *Właściwość liczbowa* przy definiowaniu zapytania. W naszym przypadku jest tylko jedna taka zmienna: *Średnica*. Aby określić parametry analizy dla tej zmiennej, rozwijamy drzewo obiektów dla naszej konfiguracji analizy (tak jak na rys. 8).

Zacznijmy od ustawienia granic tolerancji. Klikamy węzeł *Specyfikacje* i w panelu cech tego węzła (prawa część rys. 8) klikamy przycisk *Dodaj*, aby określić ustaloną wartość nominalną (docelową) dla średnicy równą 75 i szerokość przedziału dla poprawnych wartości średnicy równą 0,05 (zob. rys. 8).



Rys. 8. Specyfikacje dla karty kontrolnej.



Rys. 9. Wybór wyników analizy.

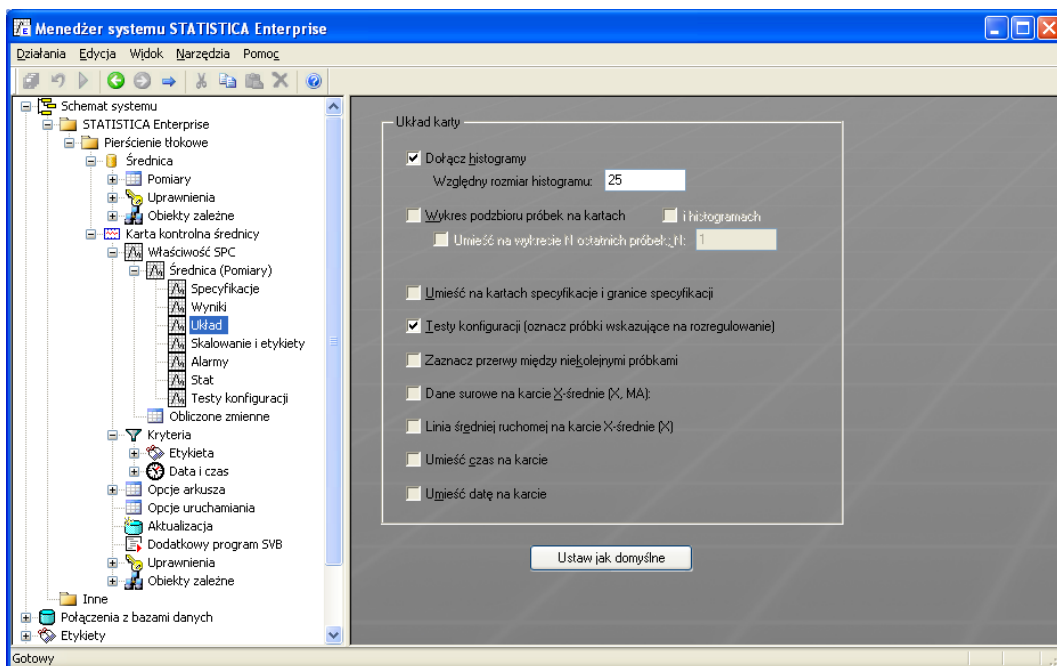


Warto zauważyć, że w ten sposób możemy definiować również wiele innych ustawień dla kart kontrolnych. Ponadto aktualne wartości specyfikacji możemy odczytywać z bazy danych, dokonując odpowiednich wpisów w polu Wyrażenie SQL dla pobierania specyfikacji (zob. rys. 8).

Następnie klikamy węzeł *Wyniki*, aby wybrać wykresy i tabele tworzone w wyniku analizy. Ustawimy tworzenie karty kontrolnej X-średnie i R oraz histogramu podsumowującego zdolność procesu (zob. rys. 9).

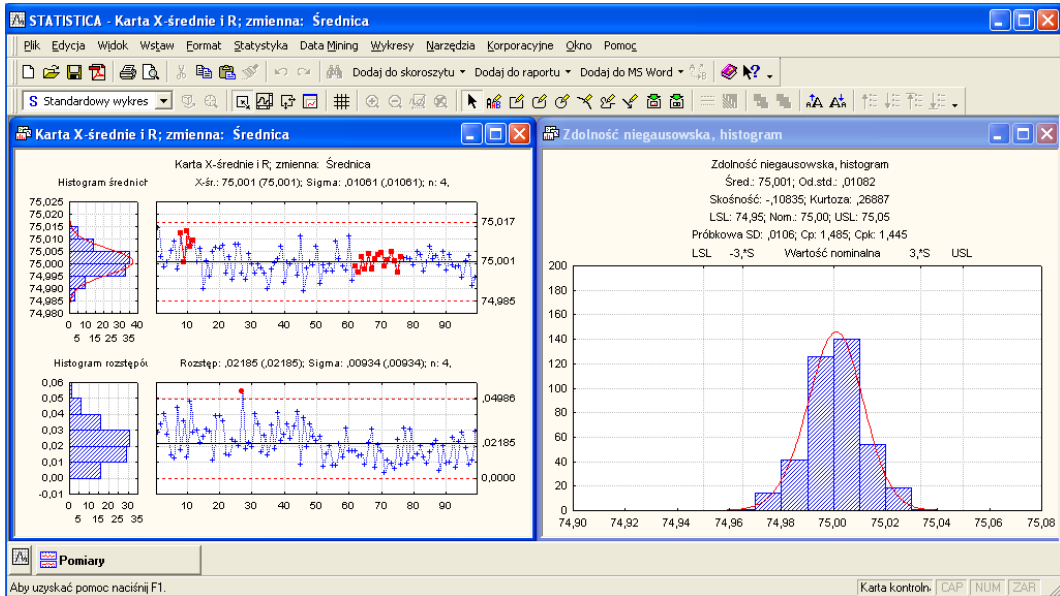
Kolejny krok określania analizy to wybór układu karty kontrolnej. Przechodzimy do węzła *Układ*. Zaznaczamy pole *Dołącz histogramy*, aby na karcie umieścić histogramy dla wartości średniej i rozstępu w próbkach. Wielkość histogramu ustawimy na 25% całego wykresu. Włączymy także sprawdzanie testów konfiguracji, aby móc wykryć nielosowy układ średnich kolejnych próbek (np. trend wzrostowy średniej procesu). Ustawienie dla układu widzimy na rys. 10.

Definiowanie analizy zakończymy, ustawiając w węźle *Data i czas* pobieranie wszystkich danych. Oprócz ustawionych przez nas opcji moglibyśmy jeszcze określić statystyki wyświetlane na karcie kontrolnej (węzeł *Statystyki*), w węźle *Alarmy* sposób reagowania na sygnały (np. próbkę poza granicami kontrolnymi) i opisywanie osi na kartach kontrolnych oraz histogramach (węzeł *Skalowanie i etykiety*). Moglibyśmy także wykonać analizę dla podzbioru danych, określając kryterium dla etykiet (np. analiza dotyczyłaby tylko jednego operatora).



Rys. 10. Układ karty kontrolnej.

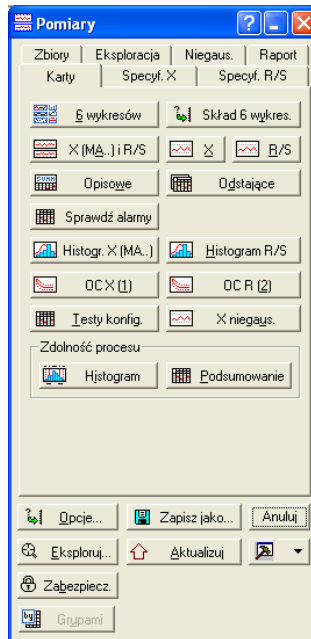
Specyfikacja analizy jest już gotowa i możemy uruchomić analizę. W schemacie systemu podświetlamy konfigurację *Karta kontrolna średnicy* i z menu podręcznego wybieramy polecenie *Uruchom*. Na ekranie pojawi się okno *STATISTICA* z wykresami, których zażądaliśmy przy określaniu analizy (rys. 11).



Rys. 11. Analiza w *STATISTICA*.

Jak widać, badany proces nie jest uregulowany: mamy dużo czerwonych punktów sygnalizujących rozregulowanie. Na karcie R mamy jedną próbkę powyżej górnej granicy kontrolnej (oznaczoną czerwonym kółkiem), a na karcie X-średnie wiele punktów wskazanych przez testy konfiguracji (czerwone kwadraty). Z drugiej strony proces ma całkiem przyzwoitą zdolność: wskaźnik  $C_p$  wynosi 1,485, a  $C_{pk}$  1,445.

Do utworzenia karty kontrolnej i histogramu podsumowującego zdolność procesu użyliśmy szablonu analizy, jednak nic nie stoi na przeszkodzie, aby analizę dalej prowadzić interakcyjnie. W naszym przypadku warto by zobaczyć, które testy konfiguracji sugerują rozregulowanie procesu – pomoże nam to zidentyfikować przyczynę lub przyczyny rozregulowania. Aby kontynuować analizę interakcyjnie, klikamy przycisk **Pomiary** na pasku na dole okna *STATISTICA*. Na ekranie pojawi się okno wyboru wyników i określania ustawień dla karty kontrolnej pokazane na rys. 12.



Rys. 12. Okno wyboru wyników i ustawień karty kontrolnej.

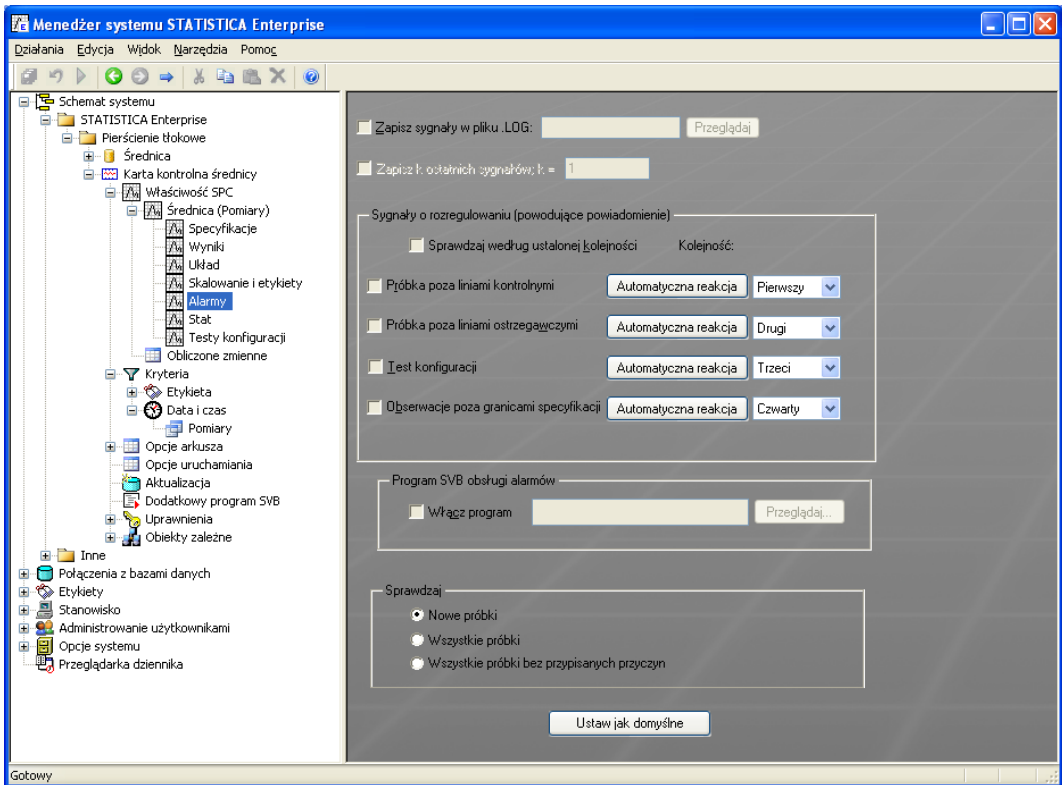
Naciskamy przycisk *Testy konfiguracji*, aby uzyskać arkusz z wynikami tych sprawdzeń przedstawiony na rys. 13.

	od próbki	do próbki
Strefy A/B/C: 3,000/2,000/1,000 *Sigma		
Testy konfiguracji		
<b>9 po tej samej stronie l. centralnej</b>	OK	OK
6 w trendzie rosnącym/malejącym	OK	OK
14 naprzemiennie w górę i w dół	OK	OK
2 z 3 w strefie A lub dalej	8	10
4 z 5 w strefie B lub dalej	8	12
15 w strefie C	62	76
8 poza strefą C	OK	OK

Rys. 13. Arkusz z testami konfiguracji.

Testy konfiguracji wykazują, że na początku mamy za dużo próbek odległych od linii centralnej (tzn. w odległych od niej strefach A i B), natomiast później z kolei próbki układają się zbyt blisko linii centralnej (w strefie C). Przypomnijmy, że 27 próbka (też w pierwszej połowie badanego okresu) jest ponad górną granicą kontrolną na karcie rozstępu (zob. rys. 11). Przyczyną wszystkich sygnałów może być trwale zmniejszenie się rozrzutu procesu, patrząc na kartę kontrolną można przypuszczać, że zmiana ta nastąpiła w okolicach próbki nr 50, co odpowiada połowie lutego 2007 r. Jeśli popatrzymy do arkusza danych, to zauważymy, że mamy kilkudniową lukę w pomiarach między 13 a 19 lutego. Ostateczne rozstrzygnięcie daje dokumentacja procesu: właśnie w tym okresie





Rys. 15. Ustawienia alarmów.

## Podsumowanie

W artykule przedstawiliśmy zasadnicze własności zintegrowanego systemu analizy danych *STATISTICA Enterprise*. Więcej informacji o tego typu systemach można znaleźć na naszej stronie internetowej w *Czytelni* w dziale *Jakość* (pod adresem [www.statsoft.pl/czytelnia/czytelnia.html](http://www.statsoft.pl/czytelnia/czytelnia.html)).

W przykładzie zaprezentowaliśmy, jak można zbudować kompletny szablon analizy, obejmujący pobranie danych, określenie ich interpretacji i samo wykonanie analizy. Wykonaliśmy kartę kontrolną, ale mogliśmy zastosować również inną metodę dostępną w *STATISTICA* (np. prognozowanie wyrównywaniem wykładniczym lub segmentację metodą k-średnich).