



CZY MÓJ PROCES JEST *TRENDY*, CZYLI ANALIZA TRENDÓW

Michał Kusy, StatSoft Polska Sp. z o.o.

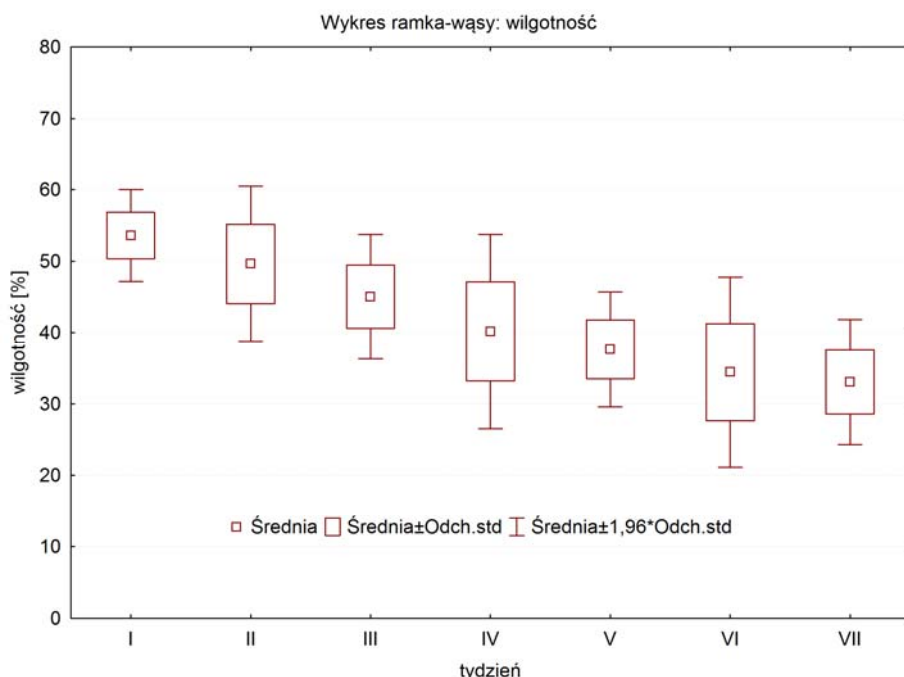
Wprowadzenie

Analiza danych w kontroli środowiska produkcji i magazynowania opiera się między innymi na szeregu metod wykrywania trendów. Pojawienie się trendu w wartościach kontrolowanego parametru może sugerować wpływ czynników zewnętrznych, które zakłócają proces produkcji (np. nieprawidłowe funkcjonowanie urządzeń wentylacyjnych, wzrost temperatury zewnętrznej itp.). Szczególnie rygorystyczne normy, jakim podlega przemysł farmaceutyczny, sprawiają, że utrata kontroli nad zachowaniem parametrów kluczowych dla procesu jest zjawiskiem wyjątkowo niekorzystnym. Zgodność z regulacjami FDA dotyczącymi dobrej praktyki wytwarzania (GMP) w farmacji wiąże się zatem z potrzebą kontroli kluczowych parametrów i przeprowadzania analizy trendów, która pozwoli ograniczyć ich wpływ na przebieg produkcji.

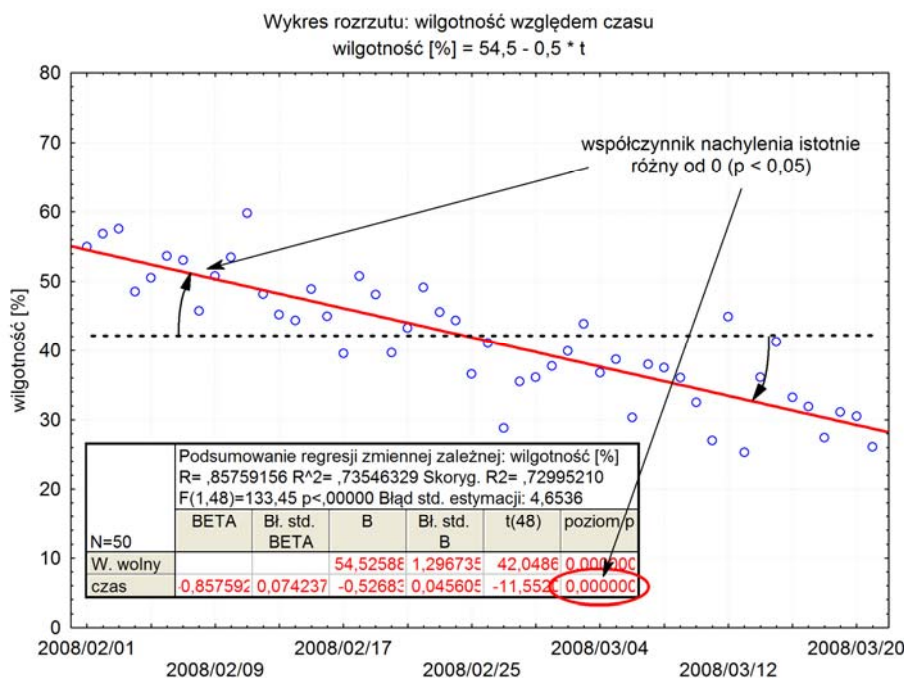
Metody wykrywania trendu

Prawidłowości występujące w zebranych pomiarach pozwala wstępnie określić analiza wykresów danych. Już na podstawie liniowego wykresu danych można zwykle zorientować się, czy pomiary sugerują występowanie jakiejś tendencji (punkty na wykresie wskazują na ogólny spadek lub wzrost wartości parametru) oraz ocenić postać funkcji trendu (tendencja ma charakter liniowy, potęgowy, wykładniczy itp.). Pomocne okazują się również skategoryzowane wykresy ramka-wąsy, na podstawie których możemy porównywać wartości parametru w wybranych przedziałach czasowych (np. dane dotyczące wilgotności w pomieszczeniu w kolejnych tygodniach lub miesiącach).

Właściwe wnioski dotyczące występowania trendu należy jednak oprzeć na odpowiednich metodach statystycznych. Występowanie trendu można badać, analizując korelacje między badanym parametrem a zmienną określającą czas. W tym celu korzystamy z testów współczynnika korelacji – parametrycznego testu współczynnika korelacji Pearsona lub nieparametrycznego testu korelacji rang Spearmana. Współczynnik korelacji przyjmuje wartości z przedziału $\langle -1; 1 \rangle$. Ujemne lub dodatnie wartości współczynnika korelacji wskazują na występowanie trendu odpowiednio: malejącego lub rosnącego. Współczynnik pozwala również określić siłę występującego trendu – im wartości są bliższe -1 lub 1, tym związek badanego parametru ze zmienną określającą czas jest silniejszy.

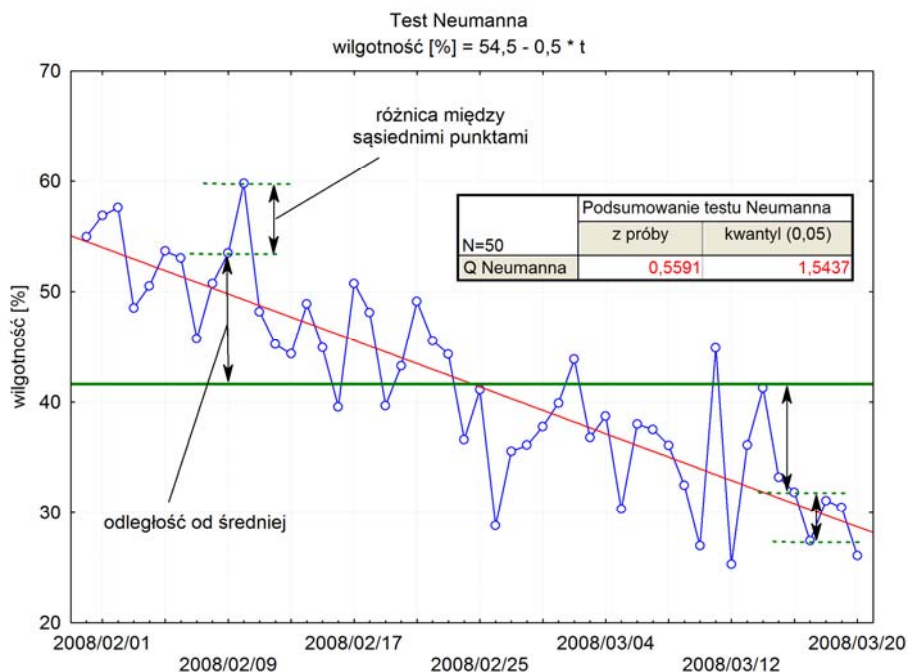


Inne, stosunkowo łatwe rozwiązanie wykorzystuje analizę regresji prostej, gdzie zmienną zależną jest badany parametr, a zmienną niezależną – czas. Wykrycie trendu sprowadza się w tym przypadku do sprawdzenia, czy współczynnik nachylenia prostej regresji jest istotnie różny od zera. Na podstawie wyliczonych współczynników regresji wyznaczane jest prawdopodobieństwo testowe, które porównujemy z przyjętym poziomem istotności α (np. $\alpha = 0,05$). W przedstawionym poniżej przykładzie wartość prawdopodobieństwa testowego $p < 0,05$ pozwala odrzucić hipotezę o braku trendu w wartościach pomiarów wilgotności. Więcej informacji dotyczących wnioskowania statystycznego znaleźć można między innymi w pozycji *Liczby nie wiedzą skąd pochodzą* [2].





Test Neumanna wykorzystywany w analizie trendu oparty jest z kolei na porównaniu różnic między sąsiednimi obserwacjami i odległości obserwacji od średniej. W przypadku trendu liniowego odległości od średniej dla początkowych i końcowych pomiarów są przeważnie większe od odległości między sąsiednimi obserwacjami. Wartość statystyki testowej w rozważanym przykładzie jest mniejsza od wartości krytycznej co pozwala na odrzucenie hipotezy o braku trendu na poziomie istotności równym 0,05.



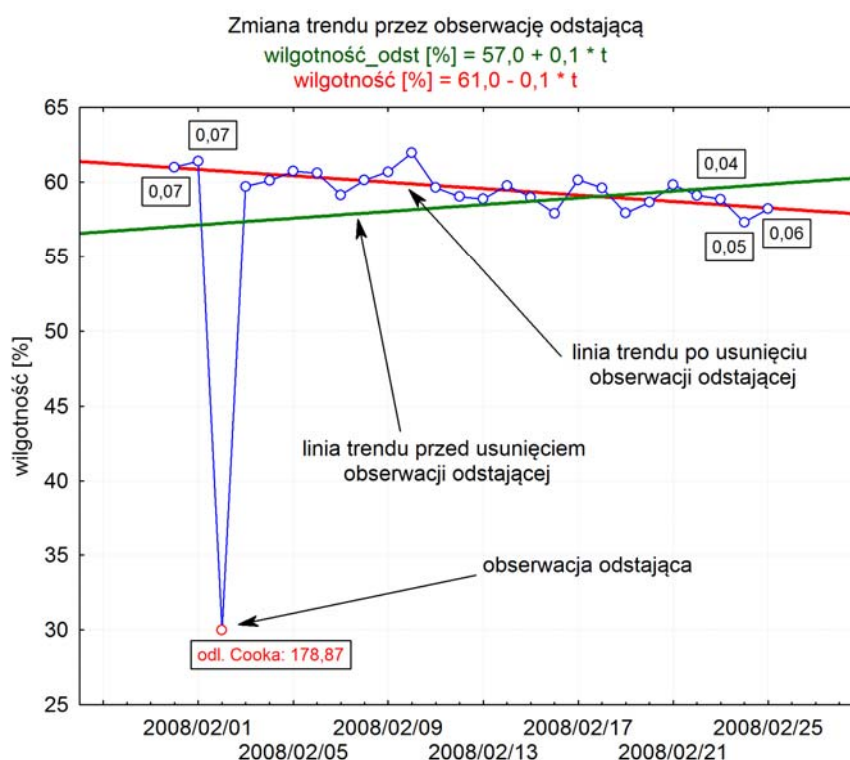
Zaletą testu Neumanna, w stosunku do testu opartego na regresji prostej są mniejsze wymagania dotyczące liczności próby. Statystyka Neumanna pozwala wykryć trendy już na podstawie kilku punktów pomiarowych. W takiej sytuacji wnioskowanie o trendzie w oparciu o dopasowanie prostej regresji jest praktycznie niemożliwe. Dodatkowo wartości parametrów regresji, podobnie jak współczynnik korelacji, wykazują się małą odpornością na działanie obserwacji odstających.

Obserwacje odstające

Przed rozpoczęciem analizy trendu należy stwierdzić, czy zgromadzone dane nie zawierają obserwacji, które zasadniczo różnią się od innych wartości i w istotny sposób zaburzają informację o rodzaju i sile występującego trendu. Obserwacje odstające mogą być wynikiem np. błędnego pomiaru lub pomyłki przy wprowadzaniu pomiaru do bazy danych. Uwzględnienie obserwacji odstających w analizie wpłynie w szczególności na nachylenie prostej regresji i będzie prowadziło do błędnych wniosków o całym procesie. Wartości, które znacznie odbiegają od pozostałych pomiarów, pozwala wykryć analiza liniowego wykresu danych. Pojawia się jednak problem, od jakiej wartości obserwacje można uznać za odstające, aby decyzja o wykluczeniu ich z analizy nie została uznana za bezpodstawną.



W przypadku analizy trendu należałoby zmierzyć, w jakim stopniu kolejne pomiary wpływają na nachylenie linii trendu.

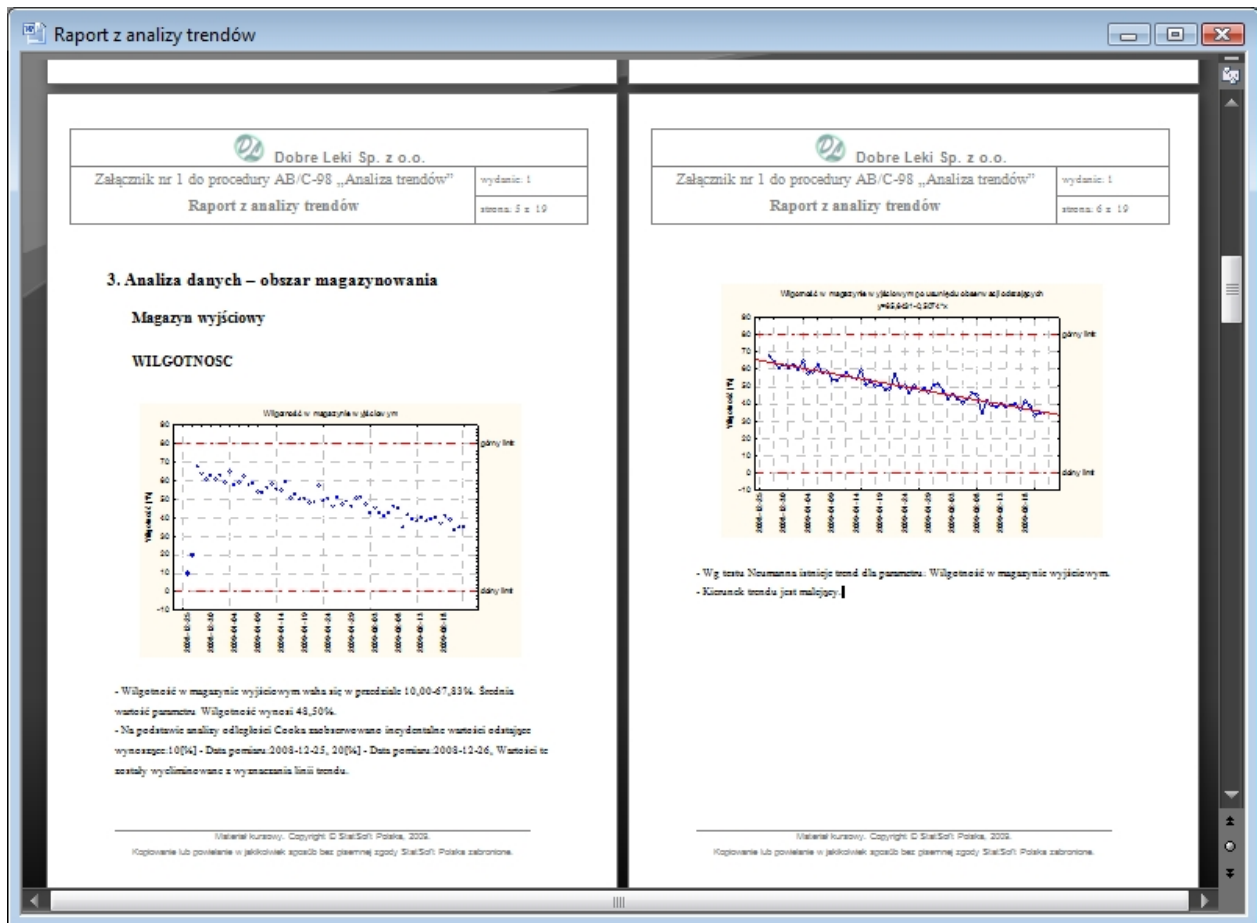


Tego typu miarą pozwalającą wykrywać obserwacje odstające jest tzw. odległość Cooka. Jest to miara wpływu danego przypadku na równanie regresji. Wskazuje ona różnice pomiędzy obliczonymi wartościami współczynników regresji a takimi samymi wartościami obliczonymi przy założeniu, że dany pomiar zostałby wyłączony z równania regresji. Jeśli odległości nie są tego samego rzędu, to można przypuszczać, że dany przypadek miał istotny wpływ na obciążenie równania regresji. Po ewentualnym wyłączeniu z badania obserwacji odstających przystępujemy do analizy występującego trendu.

Przebieg analizy trendów

Aby zwiększyć efektywność prac związanych z raportowaniem analizy trendu, StatSoft przygotował narzędzie *STATISTICA Zestaw Farmaceutyczny - Analiza Trendów*, które umożliwia automatyczne wykonywanie analizy trendów i tworzenie gotowych raportów. Program wykorzystuje odległości Cooka do wyznaczenia obserwacji odstających, dzięki czemu redukuje problemy z interpretacją tego typu pomiarów. Sama analiza trendów opiera się natomiast na powszechnie stosowanym teście Neumanna.

Analiza wykonywana przez program *STATISTICA Analiza Trendów* rozpoczyna się od sprawdzenia zakresu specyfikacji i wyznaczenia liczby pomiarów, które przekroczyły granice. Program oblicza również podstawowe statystyki próby (średnią, wartość minimalną, maksymalną itp.). Następnie, na podstawie odległości Cooka wykrywane są



Proces tworzenia raportów

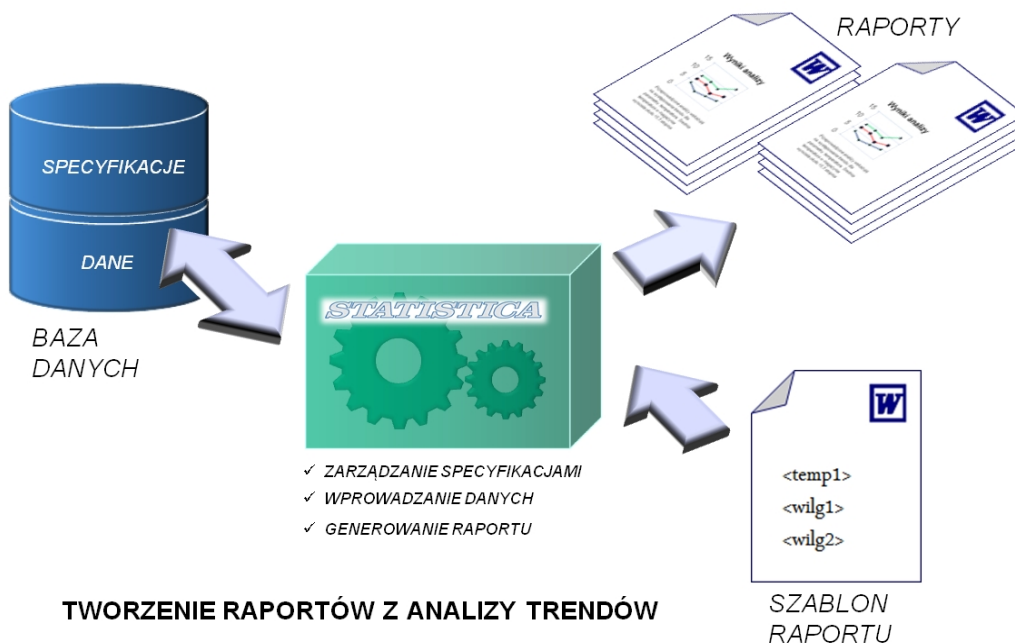
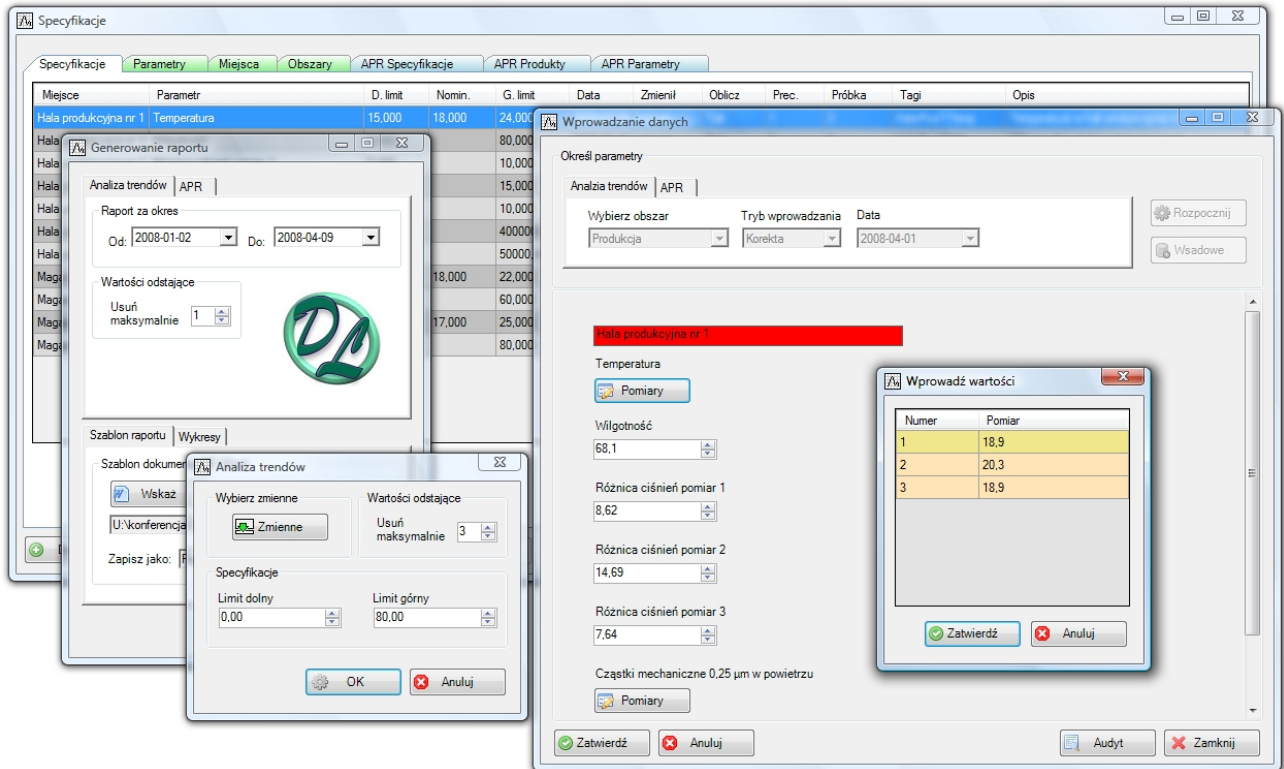
Do tej pory tego typu raporty przygotowywane były ręcznie, z wykorzystaniem analiz wykonywanych w arkuszu kalkulacyjnym. Wiązało się to z koniecznością wielokrotnego wykonywania takich samych obliczeń dla różnych parametrów i żmudnego kopiowania wyników i wykresów do dokumentu MS Word. Cały proces był bardzo czasochłonny i stwarzał wiele możliwości popełnienia błędu oraz problemów z interpretacją danych (np. które obserwacje można uznać za odstające).

Rozwiązanie *STATISTICA Analiza Trendów* umożliwiło pełną automatyzację procesu tworzenia raportów. Użytkownik projektuje szablon raportu, w którym podaje informacje, jakie miejsca i parametry powinny zostać uwzględnione w analizie (np. temperatura i wilgotność w hali produkcyjnej). Program łączy się z bazą danych, skąd, na podstawie informacji zawartych w szablonie, pobiera odpowiednie pomiary, specyfikacje i opisy. Silnik analityczny *STATISTICA* generuje wyniki analiz i wykresy, które po odpowiednim sformatowaniu są wstawiane razem z pełnymi opisami do szablonu raportu.

W ten sposób, za pomocą jednego kliknięcia, użytkownik może otrzymać raport, który tworzony był do tej pory przez kilka dni. Nasze rozwiązanie pozwala również wprowadzać do bazy danych wyniki pomiarów – zarówno pojedyncze dane, pochodzące z monito-



rowania na bieżąco parametrów procesu, jak i dane zebrane w postaci jednego pliku wsadowego. *STATISTICA Analiza Trendów* umożliwia też edycję istniejących w bazie specyfikacji i pomiarów, przy czym wszystkie zmiany są odnotowywane (razem z datą i identyfikatorem użytkownika), a istniejące dane są archiwizowane.



TWORZENIE RAPORTÓW Z ANALIZY TRENDÓW

Program *STATISTICA Analiza Trendów* pozwala dodatkowo na szybki podgląd wyników analiz, bez potrzeby generowania całego raportu. Na podstawie wskazanych przez



użytkownika zmiennych wykonywane są wszystkie opisane wcześniej analizy, a wyniki w postaci wykresów i opisów wstawiane są do skoroszytu *STATISTICA*. Dzięki zestawowi różnorodnych narzędzi statystycznych zawartemu w programie *STATISTICA* użytkownik może dodatkowo wykonać szereg innych analiz na zebranych danych.

Program *STATISTICA Zestaw Farmaceutyczny - Analiza Trendów* został wdrożony między innymi w firmie EMO-FARM Sp. z o.o., w znacznym stopniu usprawniając proces raportowania analizy trendów. Najlepszym podsumowaniem funkcjonowania naszego rozwiązania będzie opinia dyrektora ds. jakości w EMO-FARM Sp. z o.o.

„Wdrożenie nowego narzędzia umożliwiło wykonywanie odpowiednich analiz i tworzenie wymaganych raportów przy znacznym ograniczeniu zaangażowanych do tego celu zasobów. *STATISTICA Zestaw Farmaceutyczny* stanowi istotne wsparcie w zakresie dbałości o wysoką jakość leków, pozwala na znaczne oszczędności czasu, generuje czytelne i profesjonalne raporty, które umożliwiają szybką ocenę procesu, jak również są przedmiotem kontroli w trakcie inspekcji zakładu.”

Podsumowanie

Właściwe funkcjonowanie środowiska produkcji i magazynowania wiąże się z potrzebą ciągłej kontroli pewnych parametrów, w tym analizy pojawiających się trendów. Sam proces tworzenia raportów jest jednak wyjątkowo żmudny i czasochłonny. Dodatkowo stwarza wiele możliwości popełnienia błędu oraz problemy z interpretacją danych, przez co wymaga dużej uwagi w trakcie wykonywania analiz i przenoszenia wyników do raportu.

Z drugiej strony powtarzalność procesu daje możliwość jego łatwej automatyzacji. Narzędzie *STATISTICA Analiza Trendów* sprawiło, że proces tworzenia raportu sprowadza się do jednorazowego zdefiniowania specyfikacji i szablonu raportu, a następnie wybrania okresu czasu, którego raport ma dotyczyć. Oprócz oszczędności czasu i zasobów, automatyczne generowanie raportów ogranicza również możliwości pomyłek. Ponadto program ułatwia dokumentowanie zmian i archiwizowanie istniejących danych i specyfikacji.

Literatura

1. P. Dittmann, Prognozowanie w przedsiębiorstwie – Metody i ich zastosowanie, Oficyna Ekonomiczna, Kraków 2004.
2. G. Coenders, M. Saez, *Collinearity, Heteroscedasticity and Outlier Diagnostics in Regression. Do they Always Offer what they Claim?* A. Ferligoj, A. Mrvar, New Approaches in Applied Statistics, Metodološki Zvezki. 16. FDV, Ljubljana, 2000.
3. P. Francuz, R. Mackiewicz, *Liczy nie wiedzą, skąd pochodzą*, Wydawnictwo KUL, Lublin 2004.
4. K.W. Hipel, A.I. McLeod, *Time Series Modelling of Water Resources and Environmental Systems*, Elsevier, 1994.