

Czy mój proces jest trendy?

Czyli analiza trendów

Michał Kusy, Mirosław Popieluch

StatSoft Polska Sp. z o.o.

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.

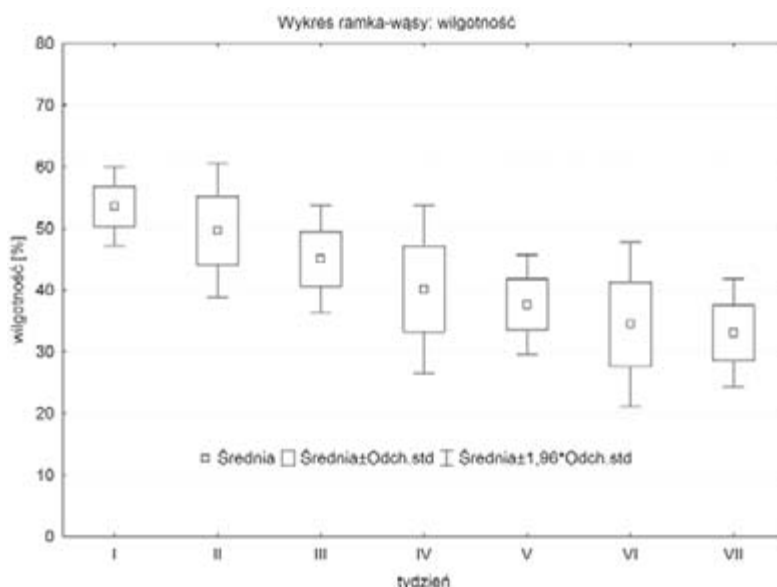
metodach statystycznych. Występowanie trendu można badać, analizując korelacje między badanym parametrem a zmienną określającą czas. W tym celu korzystamy z testów współczynnika korelacji – parametrycznego testu współczynnika korelacji Pearsona lub nieparametrycznego testu korelacji rang Spearmana. Współczynnik korelacji

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

Wykres ramka-wąsy otrzymany w programie STATISTICA



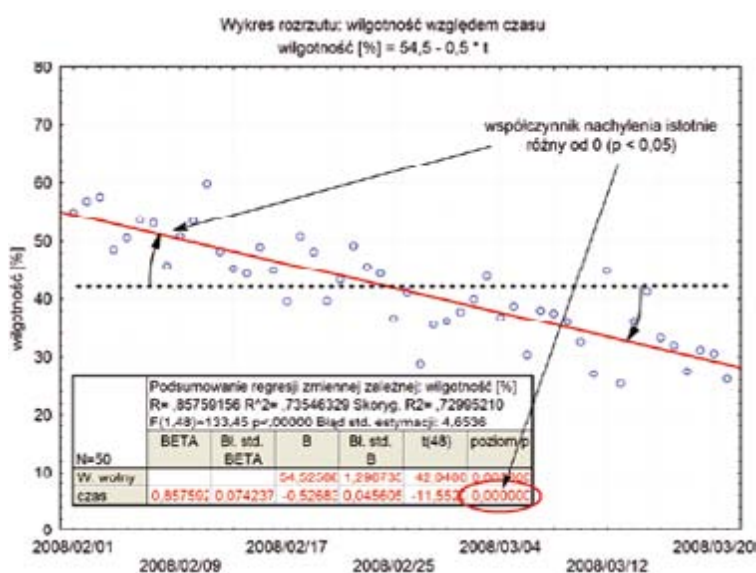
przyjmuje wartości z przedziału $<-1;1>$. Ujemne lub dodatnie wartości współczynnika korelacji wskazują na występowanie trendu odpowiednio: malejącego lub rosnącego. Współczynnik pozwala również określić siłę występującego trendu – im wartości są bliższe -1 lub 1 , tym związek badanego parametru ze zmienną 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.

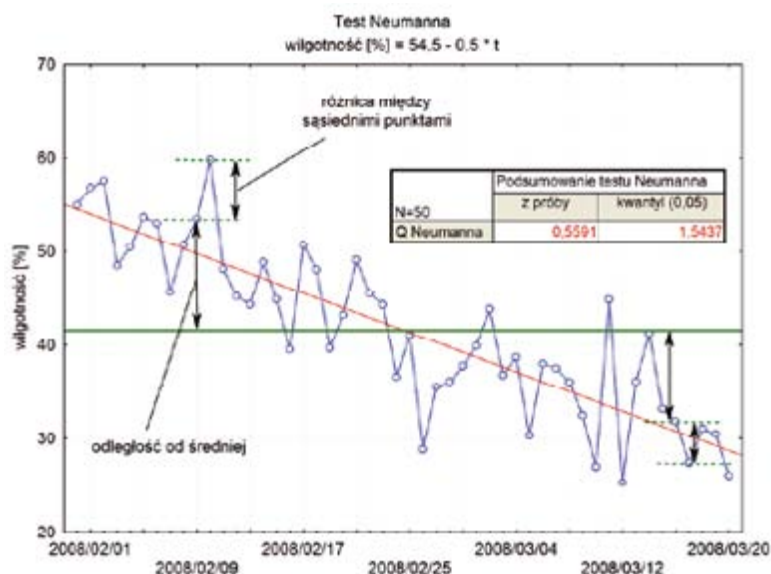
$\alpha = 0,05$). W przedstawionym poniżej przykładzie wartość prawdopodobieństwa testowego $p < 0,05$ pozwala odrzucić hipotezę o braku trendu w wartościach pomiarów wilgotności.

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 $0,05$.

Arkusz wyników i wykres otrzymany w programie STATISTICA



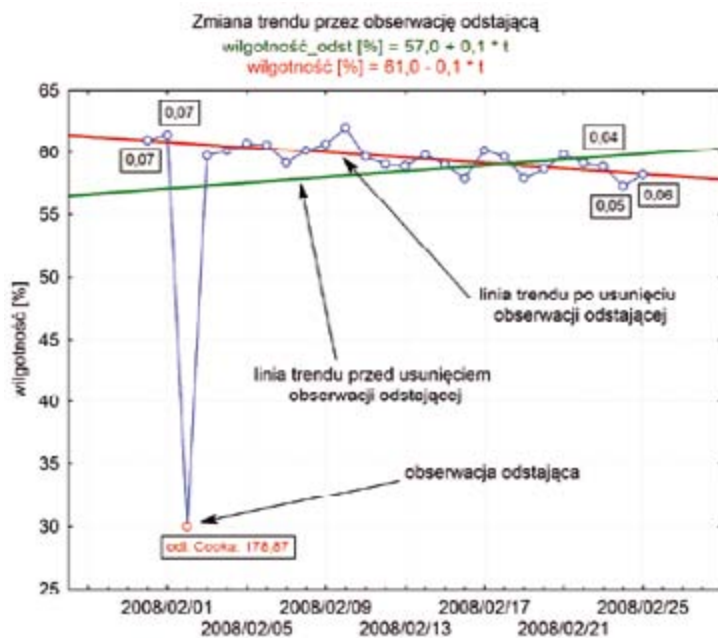
Wykres stworzony w programie STATISTICA



wydanie *.pdf do pobrania na:

www.farmacom.com.pl





Tego typu miarą pozwalającą wykrywać obserwacje odstające jest tzw. odległość Cooka. Jest to miara wpływu danego przypadku na równanie regresji. Wskazuje ona różnicę 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.

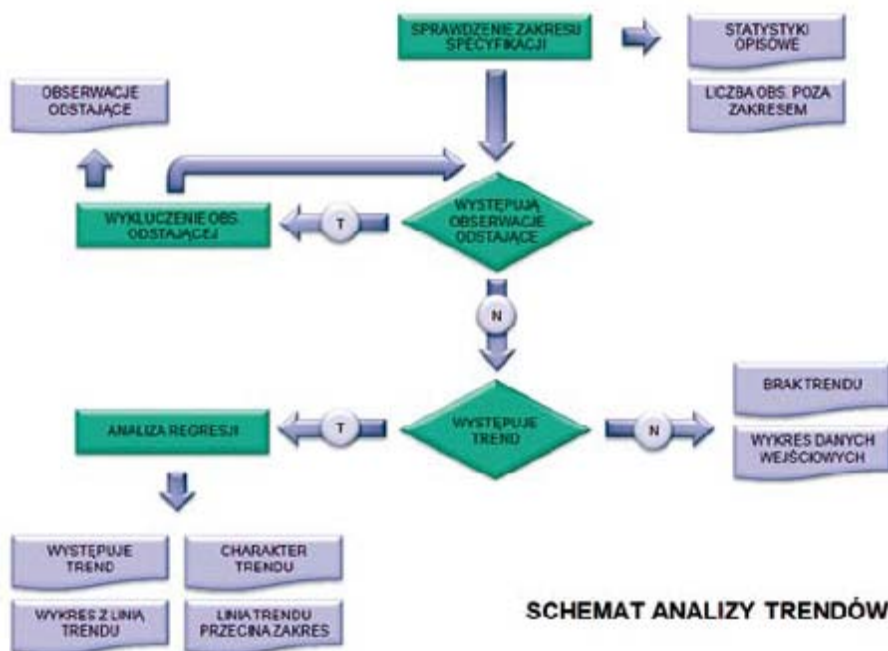
Przebieg analizy trendów

Aby zwiększyć efektywność prac związanych z raportowaniem analizy trendu, StatSoft przygotował narzędzie *STATISTICA Zestaw Farmaceutyczny - Analiza Trendów* (www.StatSoft.pl/pharma), 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 *SZF 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 obserwacje odstające. Obserwacje te zostają wykluczone z dalszych analiz, przy czym maksymalną liczbę pomijanych wartości definiuje wcześniej użytkownik. W końcowym raporcie jest uwzględniana pełna informacja o wszystkich odstających pomiarach wyłączonych z analizy.

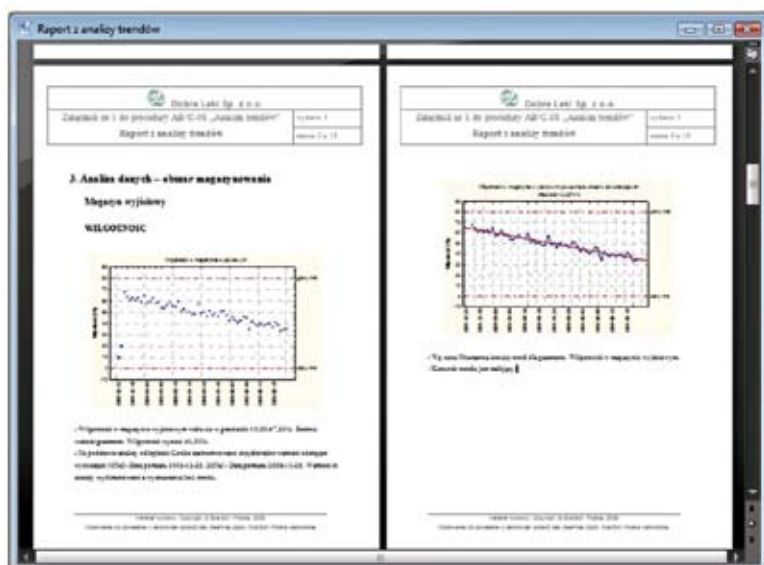
Po wyłączeniu obserwacji odstających, w oparciu o statystykę Neumanna, sprawdzana jest hipoteza o braku trendu. W przypadku wykrycia trendu program przeprowadza analizę regresji. Następnie na wykres danych nanoszona jest linia trendu. Program dostarcza również informacji o charakterze występującego trendu i sprawdza, czy linia trendu przecina granice specyfikacji. Jeśli wartości pomiarów sugerują, że w najbliższym czasie linia trendu przekroczy zakres specyfikacji, szacowany jest dodatkowo czas, po jakim granica zostanie przekroczona. Z kolei w przypadku, gdy wartość statystyki Neumanna nie wskazuje na występowanie trendu, do raportu wstawiany jest wykres danych wejściowych oraz informacja o braku trendu.



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 błędów oraz problemów z interpretacją danych (np. które obserwacje można uznać za odstające).

Fragment przykładowego raportu z programu STATISTICA Zestaw Farmaceutyczny
– Analiza Trendów



Środowisko wprowadzania danych
programy STATISTICA Zestaw Farmaceutyczny
– Analiza Trendów

Program SZF 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.

Rozwiązanie STATISTICA Zestaw Farmaceutyczny Analiza Trendów umożliwia 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 monitorowania na bieżąco parametrów procesu, jak i dane zebrane w postaci jednego pliku wsadowego. SZF 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.

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 błędów 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 Zestaw Farmaceutyczny - 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

- P. Dittmann, Prognozowanie w przedsiębiorstwie – Metody i ich zastosowanie, Oficyna Ekonomiczna, Kraków 2004.
- 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.
- P. Francuz, R. Mackiewicz, *Liczby nie wiedzą, skąd pochodzą*, Wydawnictwo KUL, Lublin 2004.
- K.W. Hipel, A.I. McLeod, *Time Series Modelling of Water Resources and Environmental Systems*, Elsevier, 1994.