

## **Case nr 3. Zaawansowana Eksploracja Danych (Specj. TPD) Szeregi czasowe i prognozowanie**

© Jerzy Stefanowski, Instytut Informatyki Politechnika Poznańska - 2011  
aktualizacja dla edycji 2013/14.

### **Cel studium przypadku:**

Studium poświęcone jest metodom analizy szeregów czasowych (ang. time series) oraz wykorzystaniu różnych metod ich dekompozycji oraz przewidywania / prognozowania przyszłych wartości szeregu. Proponowany przebieg zajęć obejmuje trzy etapy. W pierwszej części (o charakterze kilku ćwiczeń na przygotowanych prostych zestawach rzeczywistych i sztucznie wygenerowanych danych) zajmujemy się metodami adaptacyjnymi wykorzystujące tzw. mechaniczne metody wygładzania szeregów czasowych – tzn. różnego rodzaju średnie oraz wygładzanie wykładnicze. Alternatywnie pokazuje się możliwości metod analitycznych do wykrywania tzw. trendu. W drugiej części należy zapoznać się z podstawową metodą dekompozycji szeregu czasowego wykorzystującą wskaźniki sezonowości (dodatkowo można zbadać tzw. metodę Wintersa). Oprócz przykładów dydaktycznych proponuje się dokonać dekompozycji większego problemu opartego na rzeczywistych danych. Ostatnia część ma charakter typowego studium rzeczywistego przypadku – należy wybrać jeden z możliwych rzeczywistych długoterminowych zestawów danych i samodzielnie dobrać do niego najlepszy model analizy samego szeregu jak i prognozowania przyszłych wartości.

### **Pojęcia kluczowe**

Student /ka przed rozpoczęciem ćwiczenia powinna zapoznać się z następującymi pojęciami:

- Szereg czasowy i jego składniki (trend, wahania okresowe).
- Model addytywny szeregu czasowego.
- Model multiplikatywny szeregu czasowego.
- Metody średnich ruchomych.
- Wygładzanie wykładnicze oraz metoda Holta.
- Liniowe i nieliniowe analityczne funkcje trendu.
- Dekompozycja szeregu czasowego.
- Wskaźniki sezonowości.
- Metoda Wintersa
- Miary dokładności prognoz.

Powyższe pojęcia omówiono na wykładzie (patrz moja strona dydaktyczna [www.cs.put.poznan.pl/jstefanowski/tpd.html](http://www.cs.put.poznan.pl/jstefanowski/tpd.html)), tamże podano literaturę uzupełniającą. Zalecam odniesienie się do takich polskojęzycznych książek jak:

P.Dittmann: Prognozowanie w przedsiębiorstwie. Wolters Kluwer Polska, Kraków 2008.

A.D. Aczel: Statystyka w zarządzaniu (tłumaczenie). PWN, Warszawa 2000.

D.Witkowska: Podstawy ekonometrii i teorii prognozowania. Oficyna Ekonomiczna 2005.

A.Snarska: Statystyka, ekonometria, prognozowanie. Wyd. Placet 2005.

G.Box, G.Jenkins: Analiza szeregów czasowych (tłumaczenie). PWN, Warszawa 1983

P.Dittmann: Metody prognozowania sprzedaży w przedsiębiorstwie. Wyd. AE we Wrocławiu, 1983..

Warto także zapoznać się z stroną prof. K.Krawca z Politechniki Poznańskiej z materiałami dla studiów podyplomowych i niestacjonarnych – pdf obszernego wykładu „Analiza szeregów czasowych i prognozowanie”.

## Forma zaliczenia

Wykonanie ćwiczenia i zestawienie otrzymanych wyników – syntetyczny i krótki raport wyniku dla każdego z ćwiczeń najlepiej w formie elektronicznej.

## Narzędzia

Oprogramowanie Excel oraz Statsoft Statistica.

## Dane do wykonania zadań

Pliki xls zawierające dane do wykonania kolejnych zadań – patrz opisy na kolejnych zakładkach skoroszytu. Dodatkowo proste przykłady danych zawarte w tym pliku

## Przebieg ćwiczenia:

### Część pierwsze – podstawowe metody:

W pierwszej części ćwiczenia zapoznajemy się z podstawowymi metodami adaptacyjnymi, gdzie wykorzystuje się tzw. wygładzanie szeregu czasowego w oparciu o średnie (ruchome, zcentrowane, ważone itp.) lub tzw. wygładzanie wykładnicze.

### Zadanie 1. Zapoznanie się z średnimi ruchomymi.

Sprawdźmy możliwości użycia średniej ruchomej (prostej) do eliminacji losowości w przebiegu szeregu czasowego i prognozowania kolejnych wartości (zakładamy, że w tym przypadku nie ma większych wahań sezonowych albo cyklicznych, a przebieg sygnału pozwala na prognozowanie w oparciu o okno dawnych wartości zmiennej  $y$ ).

Rozważ dane z arkusza „**Kurs\_dolara**”, które przedstawiają kurs dolara w stosunku do złotówki w okresie pierwszego półrocza 2001 roku.

Celem jest sprawdzenie możliwości wygładzania tego szeregu (tj. eliminacji wahań losowych) przy pomocy średniej ruchomej prostej (np. możesz zacząć od średniej trzyokresowej  $k=3$ , a później zwiększać rozmiar okna). Następnie określ krótkoterminowe prognozy kursu na następne dni (wybrane dni po 29 06 2001). Należy także dokonać oszacowania błędu prognozy.

Wykorzystujemy model najprostszej średniej ruchomej prostej – tj. średniej z  $k$  poprzednich obserwacji. Jeśli korzystasz z Excela radzę zdefiniować formułę samodzielnie (na podstawie wiedzy z wykładu, książek i innych materiałów). W Excelu jest także funkcja ŚREDNIA RUCHOMA z dialogu *Analiza danych* dostępnych w *Narzędzia* (lecz ona ma inaczej zdefiniowane okno czasowe – wyłącznie z  $k$  obserwacją) dlatego lepiej abyś definiował formułę osobiście.

Najlepiej w kolejne wolnej kolumnie arkusza, np. C, umieścić wartości odpowiednich średnich ruchomych (oczywiście pierwsze wiersze nie mogą być obliczone z uwagi na stosowane okno czasowe). W kolejnej kolumnie D można umieścić wartości błędu między wartością prognozowaną a rzeczywistą (tzw. reszty). Na tej podstawie możesz później obliczyć globalny błąd (albo średni kwadratowy MSE, lub średni błąd przedziałowy)

Sugerowane jest wykonanie wykresów zarówno autentycznej wartości jak i prognozowanej – oceń optycznie dopasowanie wartości bieżących historycznych oraz wartości prognozowanych.

Zastanów się czy zmodyfikować wartość  $k$  oraz jak zweryfikować, które z rozważanych wartości jest najlepsza ze względu na wybrane miary oceny dopasowania i prognozowania. Na przykład użyj  $k = 4, k=5, 7$  itd. i oceń, który z parametrów lepiej przybliży rzeczywisty przebieg.

## **Zadanie 2. Dobór parametrów średniej ruchomej dla procesów przemysłowych.**

Celem tego zadania jest ponownie dobór najkorzystniejszej wartości parametru  $k$  w średniej ruchomej w oparciu o ocenę błędów prognozy.

Jako przykładowe dane rozważmy dane dotyczące obserwacji produkcji cementu w tys. ton w kolejnych miesiącach lat 1995-1997. Są one zapisane w arkuszu o nazwie „**Cement**”. Celem ćwiczenia jest ponownie znalezienie w miarę prostego modelu prognozowania – choć obecnie przebieg szeregu jest bardziej skomplikowany. Dlatego na początku wykonaj wykres produkcji w zależności od czasu i dokonaj interpretacji.

Obserwując wykres przebiegu miesięcznej produkcji cementu można zauważyć pewien regularny cykl roczny, który jest zniekształcony przez drobne nieregularności. W celu likwidacji ich wpływu zastosuj metodę średnich ruchomych, przyjmując różne stałe wygładzania  $k$ . Dla średnich centrowanych sprawdź następujące parametry:  $k = 3, 4, 5, 6$  i  $7$ .

Oczywiście w drugiej części (zadania) możesz wykorzystać te dane do budowy wskaźników sezonowości w metodach tzw. dekompozycji Cenzus I.

W celu oceny jakości prognozy dokonaj pogłębionej analizy błędów, tj. rozważ następujące inne metody oceny błędu, np.: ME – średni błąd procentowy, MAE – średni błąd bezwzględny, MSE – średni błąd kwadratowy

W oparciu o wartości tych miar podejmij decyzję, która wartość parametru  $k$  jest najlepsza.

Korzystając z najlepszego doboru parametru dokonaj prognozy produkcji miesięcznej cementu na styczeń i kilka kolejnych miesięcy 1998.

Zastanów się, czy masz koncepcję na inny model prognozy niż średnie kroczące (np. ważone, lub inne). Jeśli tak spraw ich skuteczność dla tego zbioru danych.

## **Zadanie 3. Wygładzanie wykładnicze szeregu czasowego z wykorzystaniem średniej scentrowanej.**

Celem tego zadania jest dobór najkorzystniejszej wartości parametru  $k$  w średniej ruchomej (lecz obecnie **zcentrowanej**) w szeregu, który charakteryzują silne wahania. Dobrze dobrana centrowana średnia ruchoma powinna zlikwidować te zmiany i pozwolić na wskazani funkcji **trendu zmian**. Takie podejście może przydać się w kolejnych częściach tego studium przypadku – głównie przy wyodrębnianiu składowej trendu z szeregu obdarzonego wahaniami sezonowymi w dalszej części ćwiczenia związanej z dekompozycją szeregu.

Jako przykładowe dane rozważmy dane dotyczące zapisu wielkości eksporty pewnego towaru od początku 1997 r.. Są one zapisane w arkuszu o nazwie „**Eksport**”.

Podobnie jak w poprzednim ćwiczeniu musisz dobrać wartość okna średniej centrowanej (np. zacznij od  $k = 4, 5$  itd.). Wartość formuły wprowadź w kolumnie D, odpowiednio pomijając wiersze na początku i na końcu serii danych w zależności od rozmiaru okna. Na koniec

wykonaj wykres oryginalnych wartości szeregu i wartości średniej ruchomej – optycznie oceń czy dobrze ilustrują one tendencje rozwojową eksportu towaru.

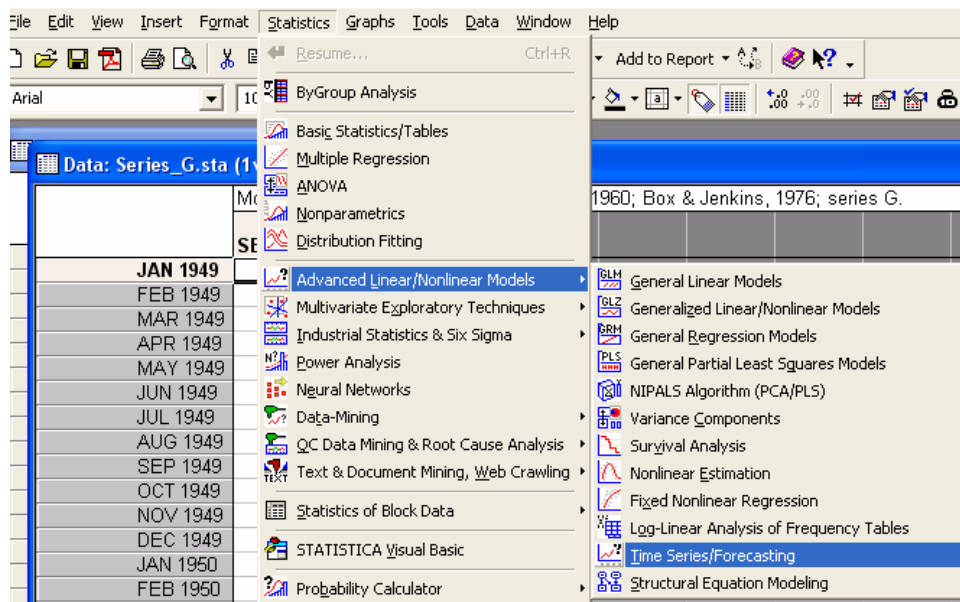
W dalszej części tego ćwiczenia rozważ alternatywne podejście analityczne do wyznaczanie linii trendu – tzn. zbuduj równanie funkcji regresji. Zdecyduj czy trend ma charakter liniowy czy nie. Wybierz w Excelu odpowiednie narzędzia modelowania funkcji regresji.

Nanieś jej przebieg na wykres – porównaj z wcześniejszym wyborem linii trendu wynikającym z średniej centrowanej.

Zdecyduj (ew. wspierając się oceną błędów), które podejście jest skuteczniejsze dla rozważanego przykładu danych eksportowych.

### Wersja ze zmienionym narzędziem – gdzie są szeregi czasowe w Statistica

W przypadku chęci skorzystania z pakietu Statistica wybierz opcje o stosowanej nazwie z menu (w obecnej wersji pakietu nazwy mogą być inne zwłaszcza w przypadku nazewnictwa polskojęzycznego)



Dalej w oknie w głównym oknie wybierz przycisk przekształcenia szeregów i dojdiesz do okna z zakładkami, gdzie może odnaleźć podopieczne wygładzania z różnymi średnimi – przykład użycia zawarto na następnym oknie.

**Przekształcenia szeregów czasowych** [?] [X]

Przekształć zmienną: SZEREG\_G: Miesięczna liczba pasażerów (w tysiącach)

Przekształcenia:  $x=f(x)$

Dodaj stałą ( $x=x+C$ ) C= -104, [▲] [▼]  
 Potęga ( $x=x^{**}C$ ) C= 2,00 [▲] [▼]  
 Pierwiastek ( $x=x^{**}1/C$ ) C= 2,00 [▲] [▼]  
 Logarytm naturalny ( $x=\ln(x)$ )  
 Przekształcenie wykładnicze ( $x=\exp(x)$ )

Odjęcie średniej ( $x=x-M$ ) M= 280,3 [▲] [▼]  
 Standaryzacja ( $x=(x-M)/S$ ) M= 280,3 [▲] [▼] S= 119,97 [▲] [▼]  
 Szacuj średnią i odchylenie standardowe z danych

Odjęcie trendu ( $x=x-(a+b*t)$ ) a= 87,653 [▲] [▼] b= 2,6572 [▲] [▼]  
 Autokor. ( $x=x-(a+b*x(\text{opóźn}))$ ) a= 0, [▲] [▼] b= 1, [▲] [▼]  
 Szacuj a/b z danych opóźn= -1 [▲] [▼]

Wyglądanie

Średnia ruch. N-punkt. N= 2 [▲] [▼]  Poprz.  Ważona  
 Mediana ruch. N-punkt. N= 2 [▲] [▼]  Poprzedzające  
 Proste wyrównywanie wykładnicze alfa= ,20 [▲] [▼]

Przekształcenia do analizy widmowej

Przekształcenia dwóch szeregów

Różnica ( $x=x-y(\text{opóź.})$ ) opóź.= 0 [▲] [▼]  
 Residualizacja ( $x=x-(a+b*y(\text{opóźn}))$ )  
 a= 0, [▲] [▼] b= 1, [▲] [▼] opóźn= 0 [▲] [▼]  
 Szacuj a i b z danych

Przesunięcie względnego początku szeregu

Przesuń szereg w przód przesun.= 1 [▲] [▼]  
 Przesuń szereg w tył przesun.= 1 [▲] [▼]

Filtrowanie i inne techniki

Filtr 4253H  
 Różnicowanie ( $x=x-x(\text{opóź})$ ) opóź.= 1 [▲] [▼]  
 Sumowanie ( $x=x+x(\text{opóź})$ ) opóź.= 1 [▲] [▼]

Wybierz przekształcenie dla zmiennej pierwotnej. Z panelu początkowego możesz wywołać wyglądanie wykładnicze.

OK (Przekształć) Anuluj

#### Zadanie 4. Wyglądanie wykładnicze szeregu czasowego.

Przy pomocy wyglądzania wykładniczego i jego modyfikacji można ekstrapolować trend (wyglądanie usuwa tzw. „szumy” i inne efekty pewnych odchyłeń, a pozostawia jedynie podstawowy sygnał), co jest przydatne do prognozowania (predykcji) zachowań szeregu w bliskiej przyszłości.

#### Klasyczna metoda Browna.

Metoda Browna (najprostsza wersja) należy do metod wyglądzania wykładniczego; stosowana jest najczęściej w przypadku szeregu bez trendu; szereg nie wykazuje tendencji rozwojowej, a wahania jego wartości wynikają z działania czynników losowych. Metoda polega na tym, że szereg czasowy zmiennej prognozowanej wyglądza się za pomocą specyficznej ważonej średniej ruchomej, przy czym wagi określone są według prawa wykładniczego.

Stosujemy podstawową formułę wyglądzania wykładniczego z uwzględnieniem elementu prognozy w momencie  $t-1$  oraz stałej wyglądzania  $\alpha$ .

Reguła predykcji w postaci rekurencyjnej:

dla pierwszego momentu czasowego:  $\hat{y}_1 = y_0$

dla kolejnych:  $\hat{y}_t = \alpha \cdot y_{t-1} + (1 - \alpha)\hat{y}_{t-1}$

Metoda prostego wyrównywania wykładniczego może służyć do prognozowania tylko na jeden krok naprzód.

Analizując ten wzór możesz zauważyć:

Jeśli wyznaczona prognoza na okres  $t-1$  była w porównaniu z rzeczywistą wartością zmiennej prognozowanej zaniżona, to prognoza na okres  $t$  zwiększa się (korekta w górę), i odwrotnie -

jeśli wyznaczona prognoza na okres była w porównaniu z rzeczywistą wartością zmiennej prognozowanej zawyżona, to prognoza konstruowana na okres zmniejsza się (korekta w dół).

Parametr  $\alpha$  nazywany jest parametrem wygładzania. Ustalenie parametru odbywa się metodą „prób i błędów”, przyjmując za kryterium, np. wartość średniego błędu prognozy ex ante. Brown uważał w swoich rozważaniach, że startowa wartości parametr  $\alpha$  może wynosić:  $2/(n+1)$ , gdzie  $n$  jest liczba obserwacji.

Uwaga rozważamy wartości  $\alpha$ , które mają być większe od zera, lecz też mniejsze niż 1!

### 1. Część – ćwiczenie doboru parametru wygładzania:

Założmy, że rozważasz następujący przykład (za L.Kowalski „Statystyka” 2003) – W pewnej hurtowni w kolejnych kwartałach lat 1998-2000 sprzedano następujące ilości żarówek (w tys sztuk:

37, 36, 34, 33, 34, 33, 35, 34, 35, 33, 34, 36

Zastosuj metodę Browna do prognozowania z tych danych, starając się dobrać najkorzystniejszą wartość parametru  $\alpha$  z uwagi na kryterium minimalizacji błędu średnio-kwadratowego (np. rozważ  $\alpha=0.2$  do  $0.8$  z krokiem co  $0.1$ ).

Wykorzystaj dowolne narzędzie (np. Excel)

### 2 Część – rozważanie danych o nieliniowym kształcie.

Można wykorzystać zarówno opcje z Analizy Danych (Excel) jak i własnoręcznie zdefiniować formuły (zalecane dla nabranie wprawy). Analogicznie radzimy wstawiać wartości wygładzane w kolejnych kolumnach.

Uwaga formuła wymaga wstawienia dla chwili  $t$  wartości rzeczywistych i prognozowanych z chwili  $t-1$  (np.  $D2=C2$ ). W przypadku pierwszej obserwacji (rok 1950) przyjmij że  $y$  prognozowane jest równe  $y$  rzeczywiste i definiuj formuły dla kolejnych momentów czasu.

W treści zadania nie podajemy wartości stała wygładzania  $\alpha$ . W trakcie ćwiczenia należy dobrać jak najlepszą wartość kierując się oceną jakości prognozy – tutaj wykorzystaj umiejętności nabyte w poprzednim zadaniu. Uwaga nie należy podnosić wartości  $\alpha$  powyżej  $0.9$ .

Przykładowe dane zawarte są w arkuszu o nazwie „**Emerytura**” – przedstawiają zmianę średniego wynagrodzenia krajowego od 1950 roku (uwzględniające normalizacje i denominacje).

Wybierz jeden z rodzajów oszacowania błędu prognozy (np. MSE lub średni przedziałowy) do podjęcia decyzji.

### Część 3 analiza alternatywnych danych

Dokonaj prognozowania wykładniczego na bardziej skomplikowanym przebiegu szeregu czasowego.

## Zadanie 5. Wygładzanie metodą Holta

Metoda ta jest udoskonaloną wersją wygładzania wykładniczego stosowaną do wygładzania szeregów, w których występują większe wahania przypadkowe. Ponadto, gdy metody wymagają dużej stałej wygładzania, a dane wykazują trend **zmieniający się** w czasie. Opis metody dostępny jest np. w pozycji „Prognozowanie i symulacje a decyzje gospodarcze”. Gajda lub „Podstawy ekonometrii i teorii prognozowania” D. Witkowska.

W ogólności wartość prognozowana w chwili  $t+1$  składa się z dwóch składników  $F$  przybliżającego poziom zmiennej oraz  $T$  oceniającego jej przyrost.

$$\hat{y}_{t+1} = F_t + T_t$$

$$F_t = \alpha \cdot y_t + (1 - \alpha) \cdot (F_{t-1} + T_{t-1})$$

$$T_t = \beta \cdot (F_t - F_{t-1}) + (1 - \beta) \cdot T_{t-1}$$

Pamiętaj, że jako wartości początkowe przyjmuje się  $F_1=y_1$  oraz  $T_1=y_2-y_1$ .

Gdy będziesz przygotowywał prognozę do wygładzonego poziomu zmiennej  $F$  dodajemy wygładzony współczynnik przyrostu trendu, ale pomnożony przez odpowiednią liczbę okresu wyprzedzenia prognozy! tj.

Równanie prognozy na okres  $t > n$  przyjmuje postać:

$$\hat{y}_t = F_n + (t - n) \cdot S_n$$

gdzie:  $y^{\wedge}$  to prognoza wyznaczone na moment  $t$  odległy od ostatniego momentu  $n$  danych z szeregu;  $F_n$  i  $S_n$  to odpowiednio wygładzone wartości zmiennej prognozowanej oraz wartości przyrostu trendu w momencie  $n$ .

Zbadaj zastosowanie tej metody do zmienionych danych z poprzedniego przykładu dotyczącego średnich rocznych wynagrodzeń – patrz arkusz „**Model Holta**”. Dobierz właściwe wartości współczynników  $\alpha$  i  $\beta$  (zaczynj np. od 0.5 i spróbuj je zmienić).

W przypadku stosowanie Excela możesz w kolejnych kolumnach umieścić wartości składnika poziomu  $F$  (np. kolumna C) i trendu  $T$  (np. kolumna D) oraz w kolejnej łączną wartość prognozy (kolumna E).

Podsumuj, która z metod jest najkorzystniejsza do prognozowania przebiegu danych.

-----

Zbadaj zastosowanie tej metody do zmienionych danych do innych danych (za Pawel Dittmann „Prognozowanie w przedsiębiorstwie”)

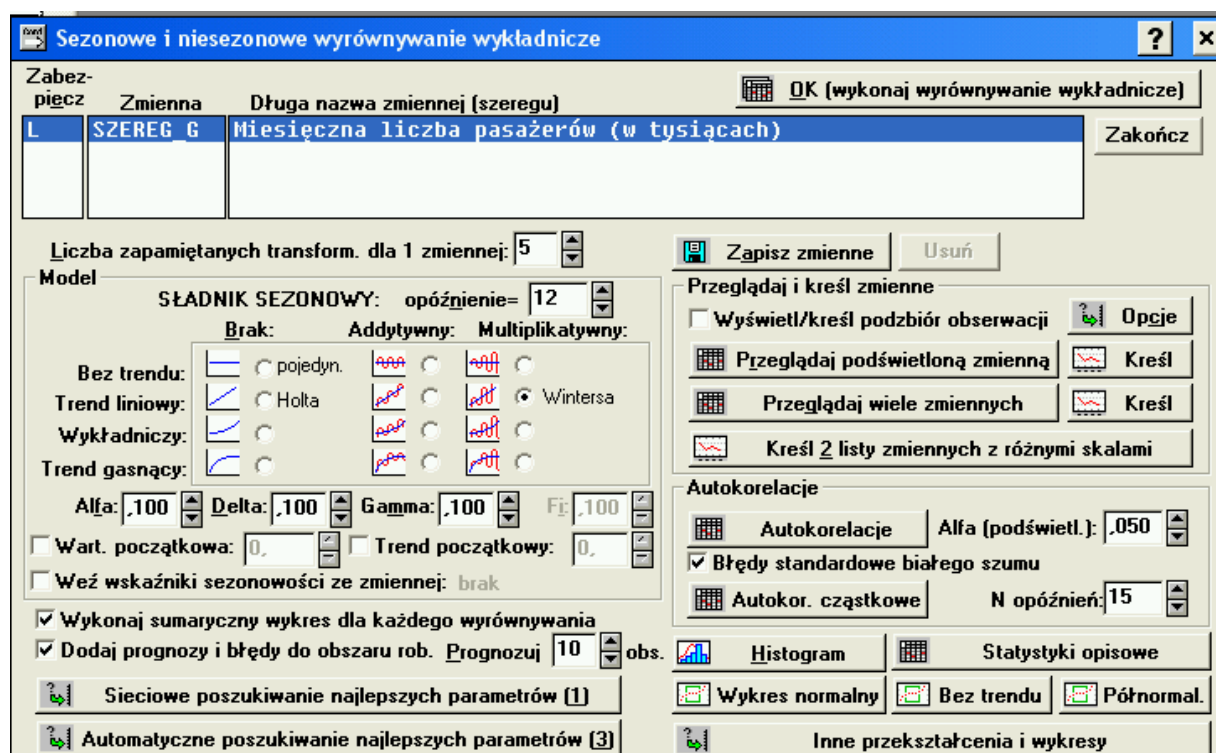
Kwartałna sprzedaż pewnego produktu np. JK kształtowała się:

370,74; 389; 390,02; 394,94; 395,85; 388,5; 391,81; 401,16; 403,73; 436,55; 388,32; 404,54; 387,05; 361; 363,98; 408,38; 435,11; 463,38; 505,32; 504,38; 557,15

Zastosuj metodę Holta (dobierając oba współczynniki  $\alpha$  i  $\beta$ ) i później dokonaj prognozy na dwa kolejne kwartały.

W przypadku zmęczenia obsługą Excela i ręcznym definiowaniem formuł użyj Statistica i odszukaj w oknie szeregi czasowe opcji Wyrównanie wykładnicze – dalej odnajdziesz opcja trend liniowy z możliwością zaznaczenia Holta (patrz zrzut ekranu za kilka stron dalej) – w mojej wersji oprogramowania spójrz na kolejność opcji *Statystyka- zaawansowane modele liniowe i nieliniowe – Szeregi czasowe i prognozowanie – Podstawowe – Wyrównanie wykładnicze i prognozowanie – Sezonowe i niesezonowe wyrównanie wykładnicze.*

Na ostatnim oknie *Sezonowe i niesezonowe wyrównanie wykładnicze* odnajdź opcje Trend liniowy i zaznacz Holta.



Zapoznaj się z opcjami 1 i 3 tzw. wspomaganie poszukiwania parametrów (Help).

## Część 2 – Analiza i dekompozycja szeregów z wahaniami sezonowymi.

### Zadanie 6. Dekompozycja szeregu czasowego z wykorzystaniem wskaźników sezonowości.

W arkuszu o nazwie „Węgiel” znajdziesz dane na temat wydobycia węgla kamiennego w latach 70tych. Wykonaj wykres tych danych, aby ustalić czy występują wahania sezonowe i jeśli tak to podejmij decyzje, jaki rodzaj modelu dekompozycji szeregu należy zastosować addytywny czy multiplikatywny?

Zastosuj metodę wskaźników sezonowości do budowy odpowiedniego modelu szeregu czasowego.

W przypadku Excela doradzamy wykrycie trendu na pomocą funkcji regresji liniowej (REGLINW).

Poniżej kilka uwag ułatwiających wykonanie ćwiczenia = proponujemy następującą organizację arkusza. W kolumnie C obok wartości oryginalnego szeregu wprowadź wartość trendu  $Y_t$ . Kolumna D może zawierać wartość  $z$  – skorygowany przebieg po eliminacji linii trendu (tj.  $C - Y_t$ ). Kolumna E – numeracja jednoimiennych momentów / faz charakterystycznych okresu – zastanów się jaką długość okresu można przyjąć (analiza wykresów oraz wiedza zdrowo-rozśądkowa) – kodujemy je indeksem  $i$ .

Kolumna F – surowe wskaźniki sezonowości  $z_i$

Kolumna G – współczynnik korygujący  $q$

Kolumna H – czyste wskaźniki sezonowości  $Q_i$

W kolejnej kolumnie umieść prognozę – tj. wybrany model szeregu (addytywny lub multiplikatywny)

Wykonaj wykres, na którym umieścisz zarówno wykres rzeczywistego przebiegi jak i jego prognozy. Wylicz także błędy prognozy.

W ostatnim kroku dokonaj prognozy wartości wydobywania węgla w roku 1979 lub dalszych.

Przećwicz swoje umiejętności na dodatkowych danych kolejnych arkuszach „Wskaźniki” oraz „Dane kwartalne”.

### Alternatywnie wykonanie tej dekompozycji w pakiecie Statistica

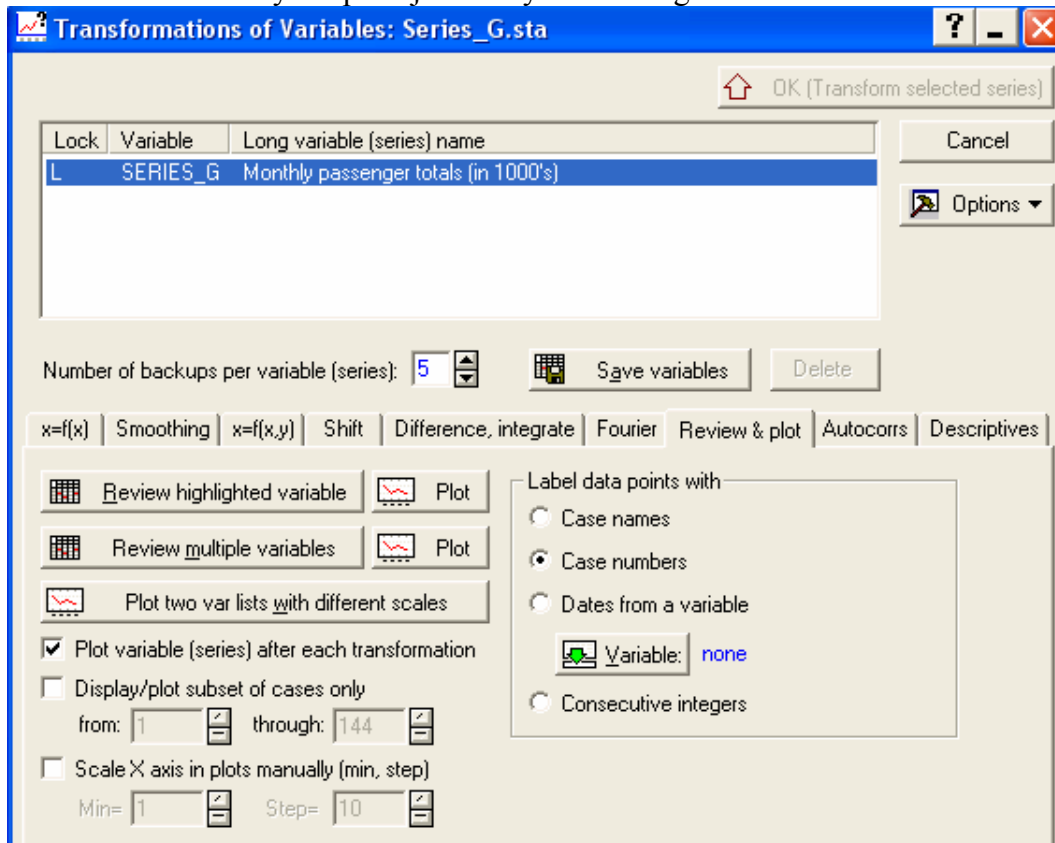
Pokazujemy poniżej przykładowe obliczenia na danych o ruchu pasażerów linii lotniczych – plik Seria G znajdujący się w standardowych przykładowych danych dla pakietu Statistica.

Wybór właściwej grupy metod w pakiecie Statistica:

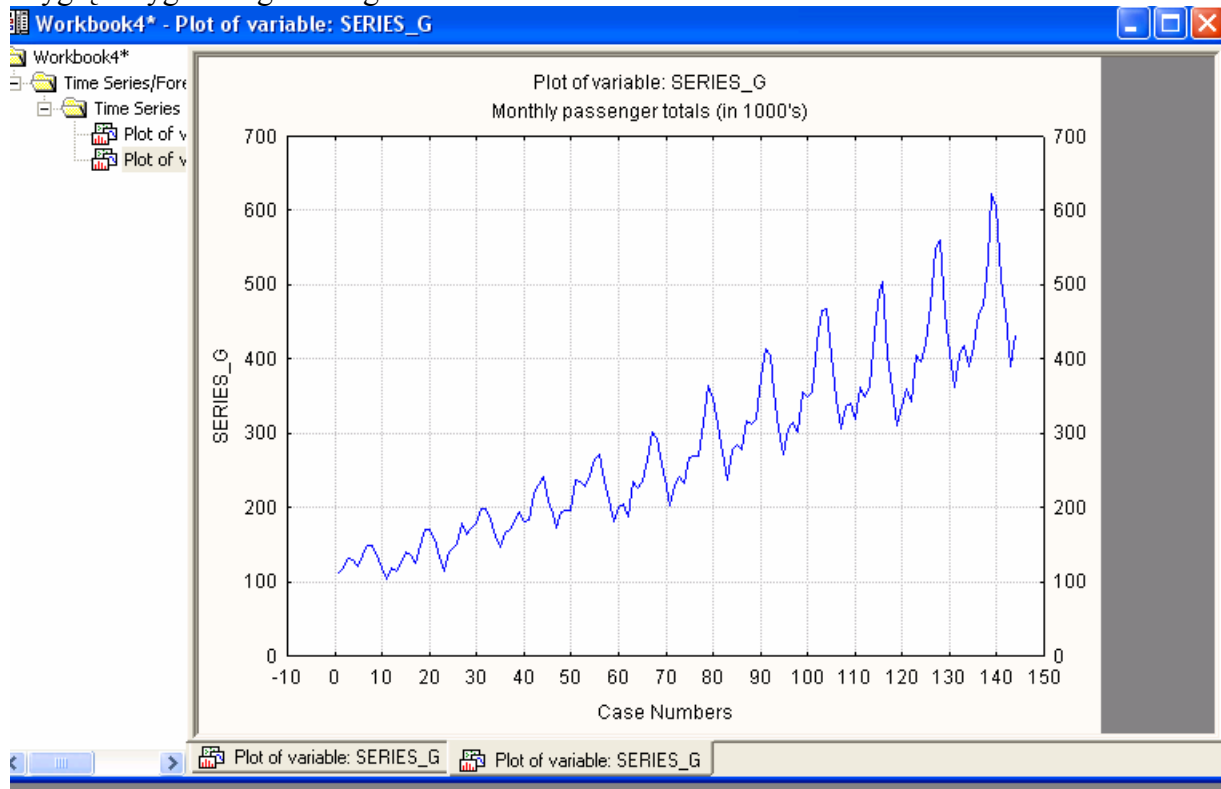
The screenshot shows the Statistica interface. On the left, a data table titled 'Monthly passenger totals (in 1000's) 1949-1960; Box & Jenkins, 1976; series G' is displayed. The table has columns for months and a column labeled 'SERIES G' with values ranging from 104 to 149. On the right, the 'Time Series Analysis: Series\_G.sta' dialog box is open. It shows the variable 'SERIES\_G' selected. The dialog includes options for 'Lock', 'Variable', and 'Long variable (series) name'. Below this, there are buttons for 'Save variables' and 'Delete highlighted variable'. At the bottom, there are several analysis method buttons: 'ARIMA & autocorrelation functions', 'Seasonal decomposition (Census 1)', 'Interrupted time series analysis', 'X11/Y2k (Census 2) - monthly', 'Exponential smoothing & forecasting', 'Distributed lags analysis', and 'Spectral (Fourier) analysis'.

Month	SERIES G
JAN 1949	112
FEB 1949	118
MAR 1949	132
APR 1949	129
MAY 1949	121
JUN 1949	135
JUL 1949	148
AUG 1949	148
SEP 1949	136
OCT 1949	119
NOV 1949	104
DEC 1949	118
JAN 1950	115
FEB 1950	126
MAR 1950	141
APR 1950	135
MAY 1950	125
JUN 1950	149

Zakładki dla możliwych operacji na danych w szeregu:



Wygląd oryginalnego szeregu



## Wybór wygładzania średnią ruchomą (k=12)

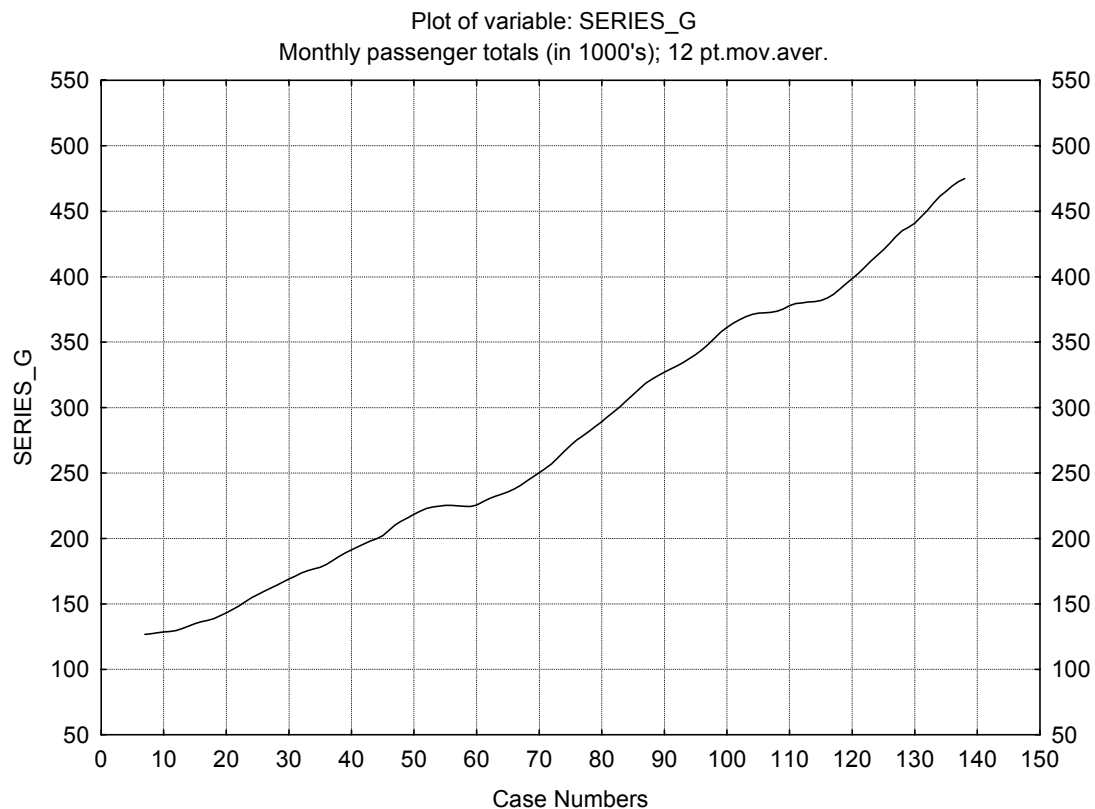
Transformations of Variables: Series\_G.sta

Lock	Variable	Long variable (series) name
L	SERIES_G	Monthly passenger totals (in 1000's)

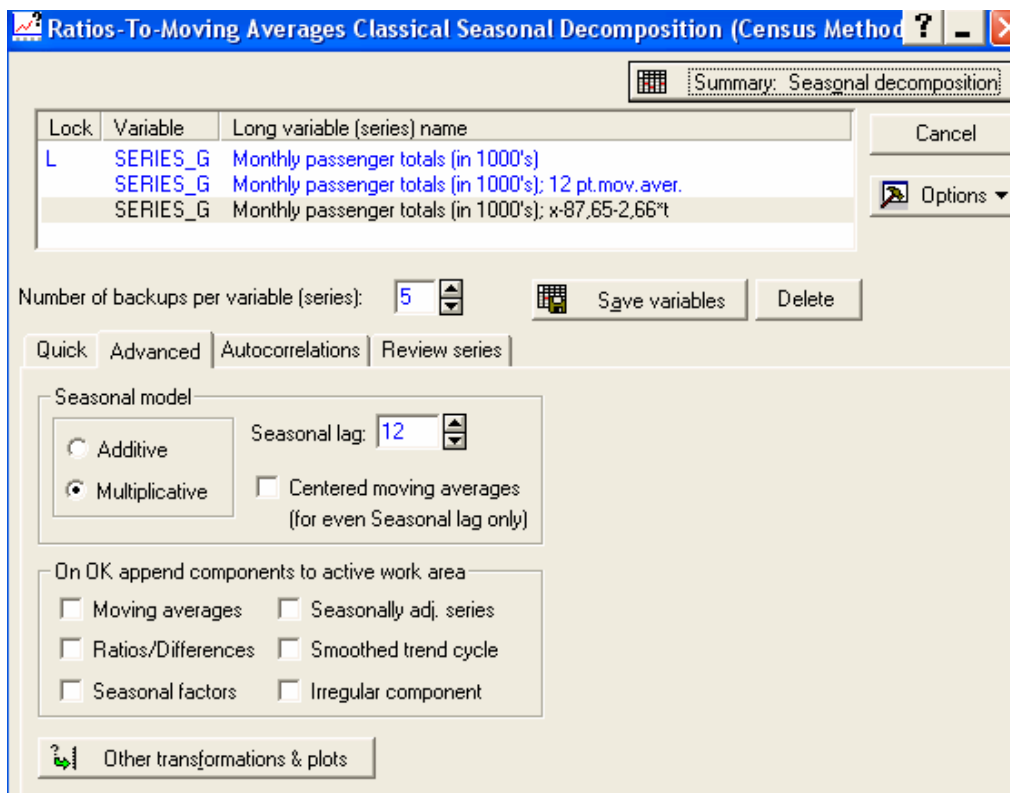
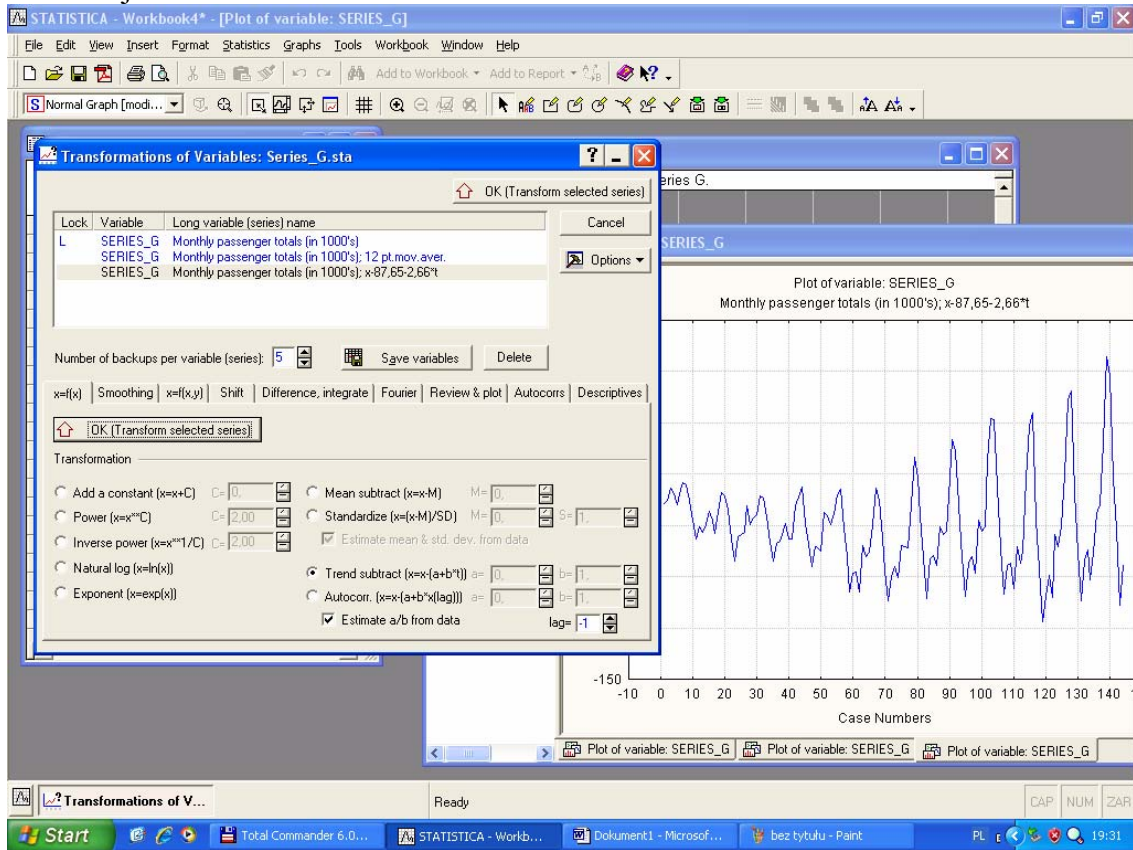
Number of backups per variable (series): 5

Transformation options:

- N-pts mov. aver. N= 2
- N-pts mov. median N= 2
- Simple exponential alpha= .20
- 4253H Filter



## Eliminacja linii trendu



Seasonal Decomposition: Multipl. season (12) (Series_G.sta)							
SERIES_G: Monthly passenger totals (in 1000's); x-87,65-2,66*t							
Case	SERIES_G trnsfrmd	Moving Averages	Ratios	Seasonal Factors	Adjusted Series	Smoothed Trend-c.	Irreg. Compon.
1	21,6900			104,0433	20,8471	27,6422	0,75
2	25,0329			160,2502	15,6211	34,7097	0,45
3	36,3757			53,7617	67,6610	48,8448	1,39
4	30,7185			50,8392	60,4228	53,1880	1,14
5	20,0613			31,5007	63,6853	48,6417	1,31
6	31,4041			164,3331	19,1100	33,7792	0,57
7	41,7469	21,7422	192,0	208,8093	19,9928	26,3518	0,76
8	39,0898	19,3350	202,2	206,7856	18,9035	21,6110	0,87
9	24,4326	17,3445	140,9	65,1963	37,4754	21,3868	1,75
10	4,7754	15,4373	30,9	127,6063	3,7423	14,8082	0,25
11	-12,8818	13,2801	-97,0	-87,2100	14,7710	8,8509	1,67
12	-1,5390	10,9563	-14,0	114,0842	-1,3490	1,7911	-0,75
13	-7,1962	9,4658	-76,0	104,0433	-6,9165	1,9829	-3,49
14	1,1466	8,6419	13,3	160,2502	0,7155	5,1836	0,14
15	13,4895	7,8181	172,5	53,7617	25,0912	7,1064	3,53
16	4,8323	6,9942	69,1	50,8392	9,5050	4,2176	2,25
17	-7,8249	5,5037	-142,2	31,5007	-24,8404	0,1434	-173,28
18	13,5179	3,6798	367,4	164,3331	8,2259	3,2379	2,54
19	31,8607	2,8560	1115,6	208,8093	15,2583	9,7715	1,56
20	29,2035	2,6988	1082,1	206,7856	14,1226	12,8288	1,10
21	14,5464	2,0416	712,5	65,1963	22,3116	14,4174	1,55
22	-13,1108	2,4678	-531,3	127,6063	-10,2744	10,8491	-0,95
23	-34,7680	2,1439	-1621,7	-87,2100	39,8670	10,2894	3,87
24	-11,4252	3,4034	-335,7	114,0842	-10,0147	1,9723	-5,08
25	-9,0824	3,1629	-287,2	104,0433	-8,7294	2,2046	-3,96
26	-6,7396	2,9224	-230,6	160,2502	-4,2056	3,4419	-1,22
27	18,6033	2,6818	693,7	53,7617	34,6032	12,6144	2,74

Raport składników dekomponowanego szeregu i prognozy

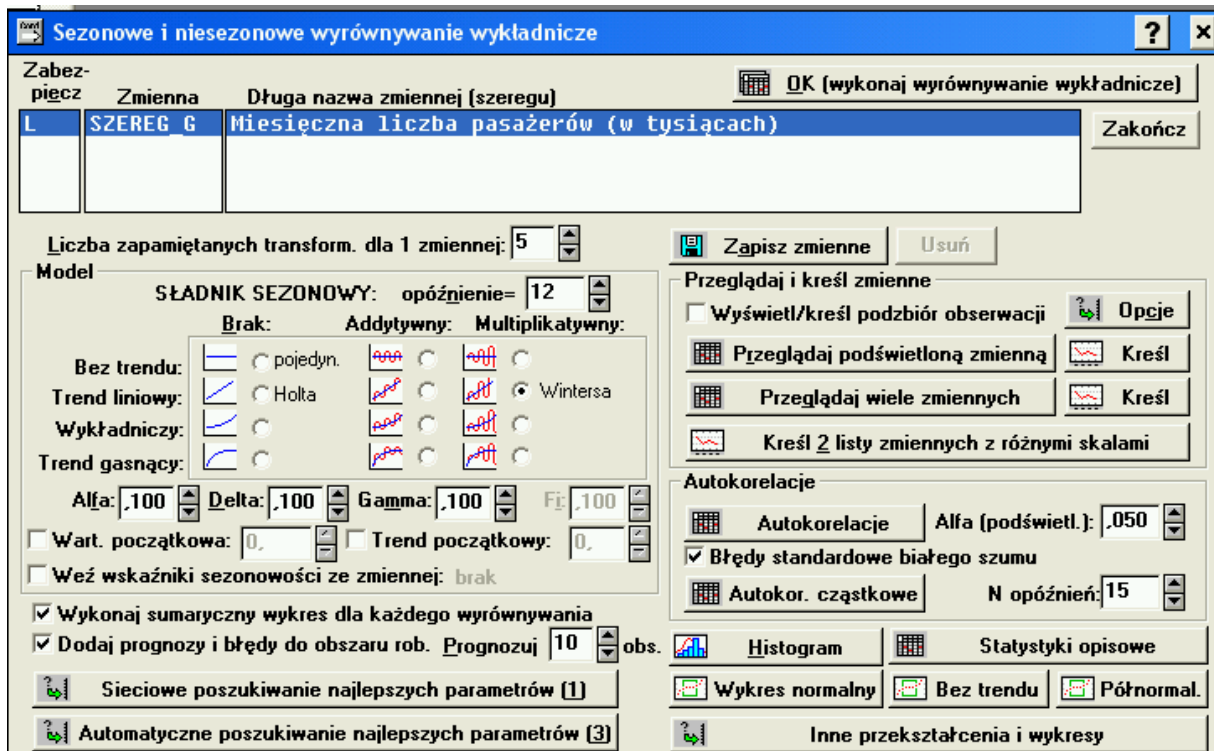
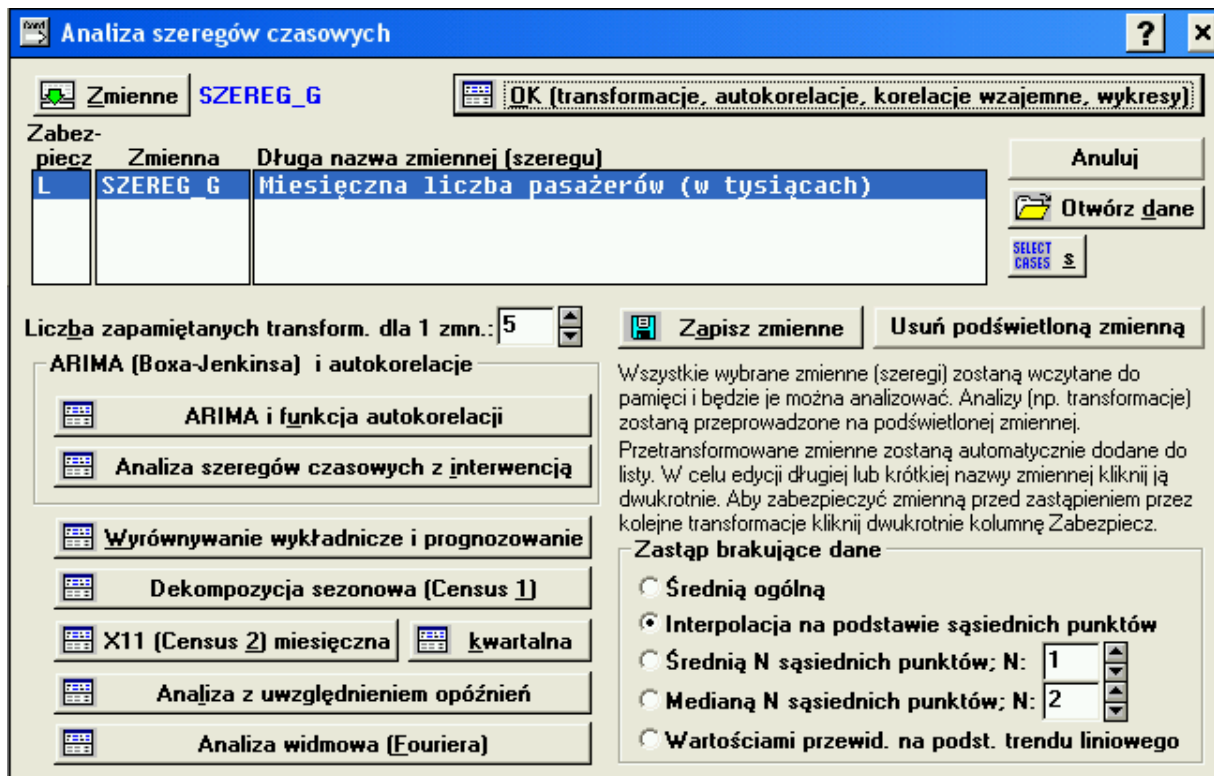
## Zadanie 6. Prognoza dla szeregu czasowego metodą Wintersa.

Opis metody Wintersa jest podany na wykładzie + opisany w części podręczników. Przypomnijmy, że jest to rodzaj uogólnienia metody Holta na przypadek szeregu z wahaniami sezonowymi. W obliczeniach formuł uwzględnia się dodatkowy składnik S z dodatkowym parametrem  $\gamma$ .

Dla obliczeń sugerujemy wykorzystanie arkusza **SezonowoscWinters** i danych tam zawartych.

Dla oceny skuteczności prognoz nie stosuj wyłącznie oszacowań błędów, lecz także wykonaj wykresy obserwując na ile prognozowany sygnał podąża za właściwymi wartościami danych.

Jeśli nie chcesz samodzielnie budować formuł w Excelu to możesz także skorzystać z opcji Statistica – w pierwszym z okien wyboru poszukaj opcji Wyrównanie wykładnicze i prognozowanie – po jej wybraniu zobaczysz specjalne okno zatytułowane Sezonowe i niesezonowe wyrównanie wykładnicze – spójrz na kolejny rysunek. Możesz tam wybrać kolejną opcję model Wintersa zarówno dla wersji addytywnej jak i multiplikatywnej.



W celu lepszego poszukiwania parametrów możesz skorzystać z metod poszukiwania najlepszych parametrów.

## **Dodatkowe część → dekompozycja rzeczywistego szeregu czasowego z nieliniowym trendem**

W celu zapoznania się z bardziej rzeczywistym przykładem – rozważ rzeczywiste dane, o większej wielkości oraz bardziej skomplikowanym, nieliniowym trendzie.

Przykładowo wczytaj dodatkowe dane z pliku o nazwie beerproduction.xls (zawierają one dane o miesięcznej produkcji piwa (w megalitrach) w Australii w latach 1956-1995 (źródło ABS).

Wykonaj najpierw wykres zmian y od czasu, aby ocenić zjawisko i dobrać właściwe narzędzie (dekompozycja, wygładzania wykładnicze albo inne). Staraj się znaleźć tą metodę, która minimalizuje stosowany błąd.

## **Trzecia część → Zaawansowana analiza i prognozowanie z rzeczywistego szeregu czasowego**

Ta część zajęć odpowiada samodzielnemu studium przypadku. Należy odnaleźć na jednym ze wskazanych serwisu kolekcje rzeczywistych zbiorów danych typu szeregi czasowe. W przypadku niektórych z nich są dodatkowe opisy podające ich źródło, charakterystykę itp.

Sugeruje wybór jednego z odnalezionych zbiorów danych – możliwie zbliżonego do dotychczas analizowanych, tzn. byłoby dobrze, aby przebieg danych nie był mocno zaburzony czynnikami „losowymi”, nie występowały obserwacje odstające lub silne nieregularności przebiegu (ocena optyczna na podstawie wykresu) i nadawał się do analizy poznanymi metodami. Rozmiar danych – to jest liczba obserwacji powinna być większa niż w dotychczas badanych zadaniach.

W pierwszym etapie należy doprowadzić dane do formatu aprobowanego przez wybrane narzędzie eksploracji danych.

Warto w drugim etapie wykonać wykres przebiegu szeregu, aby zorientować się w ich charakterystyce.

Głównym celem tego ćwiczenia jest samodzielny wybór najlepszej metody prognozowania wartości w tym problemie. Można użyć kilku metod i na podstawie błędu prognozy wskazać najlepszą.

Sugerowane internetowe repozytoria danych dotyczących szeregów czasowych:

1. Time Series Data Library przygotowane przez Rob'a Hyndmana (duża kolekcja ponad 800 zbiorów danych z różnych dziedzin zastosowań)

<http://robjhyndman.com/TSDL/>

2. Zestaw danych przygotowany przez Petera Dunna (zawiera także odnośniki do innych repozytoriów publicznie dostępnych danych statystycznych)

<http://www.sci.usq.edu.au/staff/dunn/Datasets/tech-timeseries.html>

3. Ponad 10 różnych danych z Duke University w US

[http://www.stat.duke.edu/~mw/ts\\_data\\_sets.html](http://www.stat.duke.edu/~mw/ts_data_sets.html)

4. Prosty zestaw danych z Iowa Univerisity (głównie związane z książką Time Series Modelling of Water Resources and Environmental Systems by K.W. Hipel and A. I. McLeod; lecz także odnośniki do innych zasobów amerykańskich)  
<http://www.stats.uwo.ca/faculty/aim/epubs/datasets/default.htm>

5. Economagic.com: Economic Time Series Page – amerykańska kolekcja odnośników do danych ekonomicznych, biznesowych, itp.

6. Analiza raportów nt. kwartalnej sprzedaży firmy JCPenney (np. sprzedaż od 1996 do 2001 jest często wykorzystywana w angielskich podręcznikach nt. szeregów czasowych). Raporty finansowe są publiczne dostępne (po odpowiednim sformatowaniu na stronie firmy [www.jcpenney.net/company/finance/finarch.htm](http://www.jcpenney.net/company/finance/finarch.htm))

Możesz też poszukać innych repozytoriów danych.