Analiza Szeregów Czasowych

Plan

1. **Uwagi wstępne**
   (szeregi, przykłady, prognozowanie,…)
2. **Cel analizy szeregów czasowych**
3. **Struktura szeregów czasowych**
   (trend/składowa stała, wahania sezonowe, wahania cykliczne, wahania przypadkowe)
4. **Podstawowe modele matematyczne**
   – addytywne
   – multiplikatywne
5. **Szereg czasowy bez trendu**
6. **Szereg czasowy z trendem (wyodrębnianie trendu)**
7. **Analiza sezonowości**
8. **Zagadnienia prognozowania**

@Jerzy Stefanowski, Inst.. Informatyki PP

Poznań 2002/3, aktualizacja 2009 dla TPD - ZED

Wprowadzenie

- *Time-Series Data*, *Time-related data* – dane zmieniające się wraz z upływem czasu; dane zawierające serie (szeregi) wartości / wielkości zmieniających się w czasie.
- **Szereg czasowy** – ciąg obserwacji pewnego zjawiska w kolejnych jednostkach czasu [def. statystyczna].
- Wielkości mierzone na skalach liczbowych, na ogól w równych odstępach czasu; Kształt wykresu niesie istotną informacje.

  Ogólna postać szeregu czasowego

Czas (t)	Zjawisko (y_t)
t_1	y_1
t_2	y_2
.	.
.	.
.	.
t_n	y_n

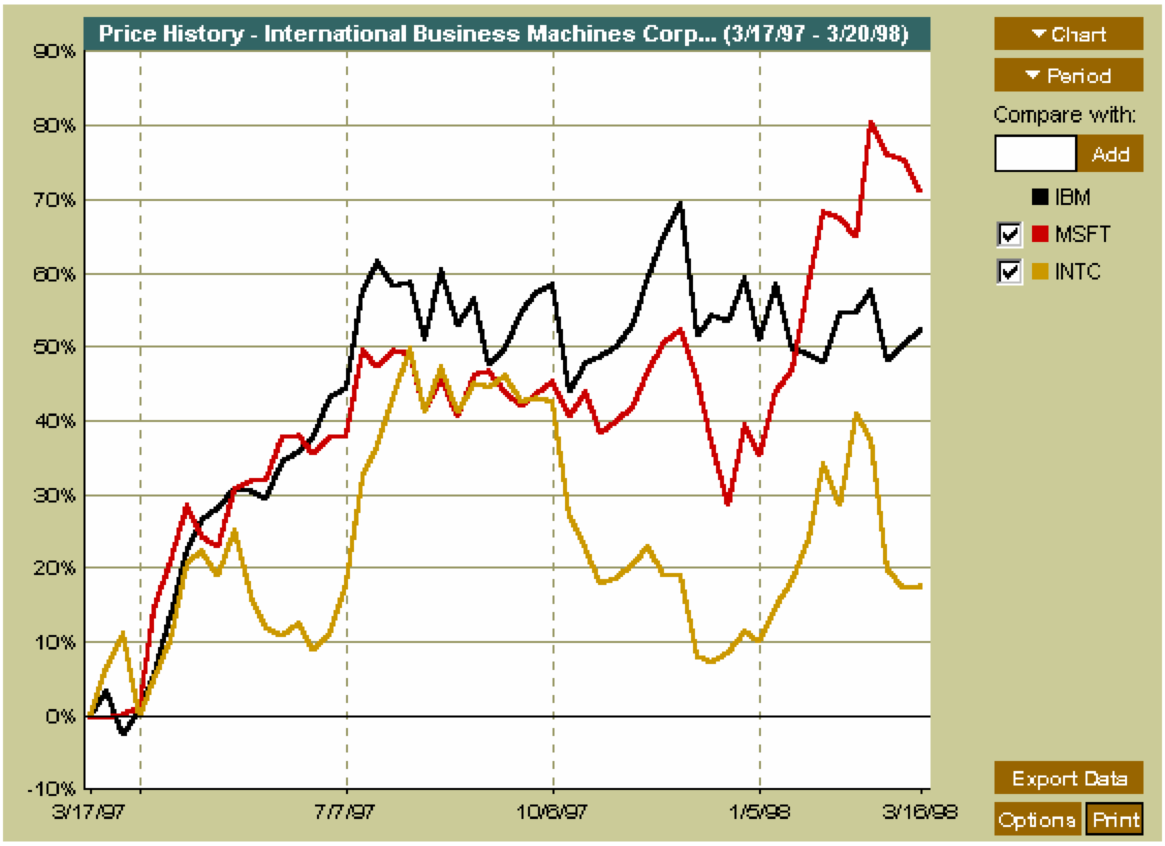
- Szeregi czasowe są podstawą analizy **dynamiki** zjawisk.

  Metody indeksowe (popularne w zastosowaniach ekonomicznych).

  Identyfikacja struktury szeregu czasowego.

Przykłady szeregów czasowych

Date	Stock	Price$
June 11, 93	IBM	98.5
June 11, 93	MSFT	78.0
June 11, 93	INTC	76.5
June 12, 93	IBM	99.5
June 12, 93	MFST	80.0
June 12, 93	INTC	77.0
June 13, 93	IBM	98.0
:	:	:
:	:	:



Szeregi czasowe

- Dane reprezentowane w postaci szeregów czasowych są popularne w wielu zastosowaniach, np.:
  - Analiza danych giełdowych.
  - Opracowywanie danych GUS (spójrz *Roczniki Statystyczne* lub *Biuletyny Statystyczne*).
  - Wspomaganie decyzji w zarządzaniu przedsiębiorstwami, w szczególności tworzenie **prognozy sprzedaży** a także analiza **dynamiki** procesów produkcyjnych, zaopatrzenia, zapasów, finansów, siły roboczej.
  - Analiza danych diagnostycznych i prognozy postępowania w medycynie.
  - Analiza wyników eksperymentów naukowych.
  - **…**
- Analiza szeregów czasowych (ang. *time series*) jest powiązana z metodami **prognozowania** (ang. *forecasting*).
- Należy odróżnić *time-series data* od *sequence data*.

Cel analizy szeregów czasowych

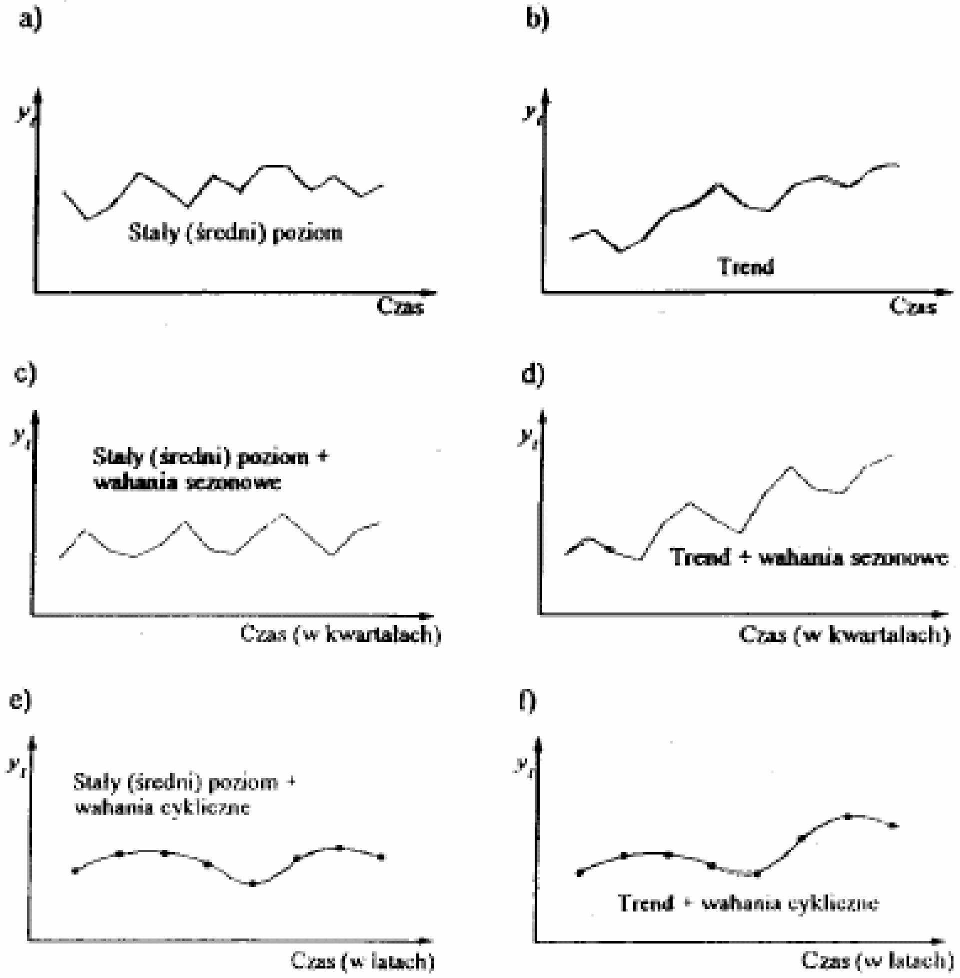
- Zbudowanie modelu pewnego zjawiska/procesu w oparciu o obserwowane zmiany w czasie pewnych mierzalnych wielkości opisujących ten proces.
- Ogólne założenie: obserwowany przebieg składa się z:
  - **Części systematycznej** (trend, składowa stała, wahania sezonowe i cykliczne) – w oparciu, o które buduje się model.
  - **Części przypadkowej** (szumu, wahań przypadkowych).
- Wymienione składniki – czynniki determinujące rozważane zjawisko; W analizie szeregów dąży się do ich wyodrębnienia i pomiaru – **dekompozycja szeregu czasowego.**
- Przy użyciu otrzymanego modelu można dokonywać predykcji (eksploracji) przebiegu szeregu lub jego składowych.

Podstawowa struktura szeregów czasowych

- **Stały** (przeciętny) poziom zmiennej.
- **Trend** (tendencja rozwojowa) – reprezentuje ogólny kierunek rozwoju zjawiska (systematyczne zmiany, jakim podlega zjawisko); rozróżnia się, np., trend liniowy lub nieliniowy.
- **Składowa okresowa** (wahania okresowe / regularne odchylenia od tendencji rozwojowej) – składnik powtarzający się cyklicznie.
- **Szum** (zakłócenia, wahania przypadkowe).

Składowa okresowa może wystąpić w postaci wahań:

- **cyklicznych** - długookresowe, rytmiczne wahania (cykl koniunkturalny gospodarki, cykl rozwoju populacji nabywców danego produktu, itp.),
- **sezonowych** – krótkookresowe do 1 roku, odzwierciedlają wpływ zachowań wynikający z „kalendarza” (np. rytm pracy w skali tygodnia, dnia, pory roku, świąt, ...).



Rys. 4.2. Szeregi czasowe z różnymi rodzajami składowych

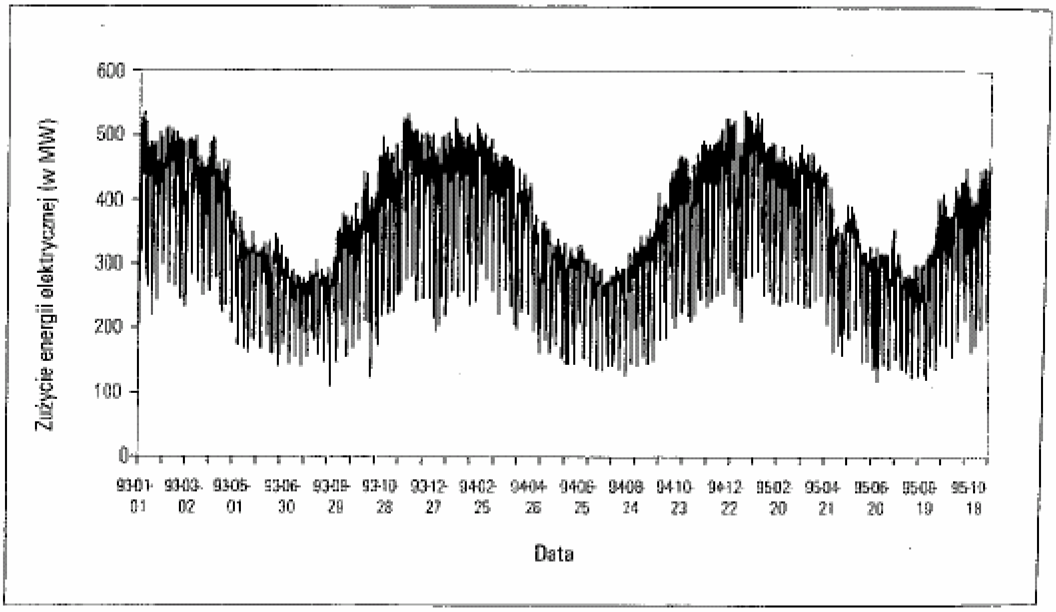
Przykład – sprzedaż energii elektrycznej

Źródłó D.Witkowska (Podstawy ekonometrii i teorii prognozowania)

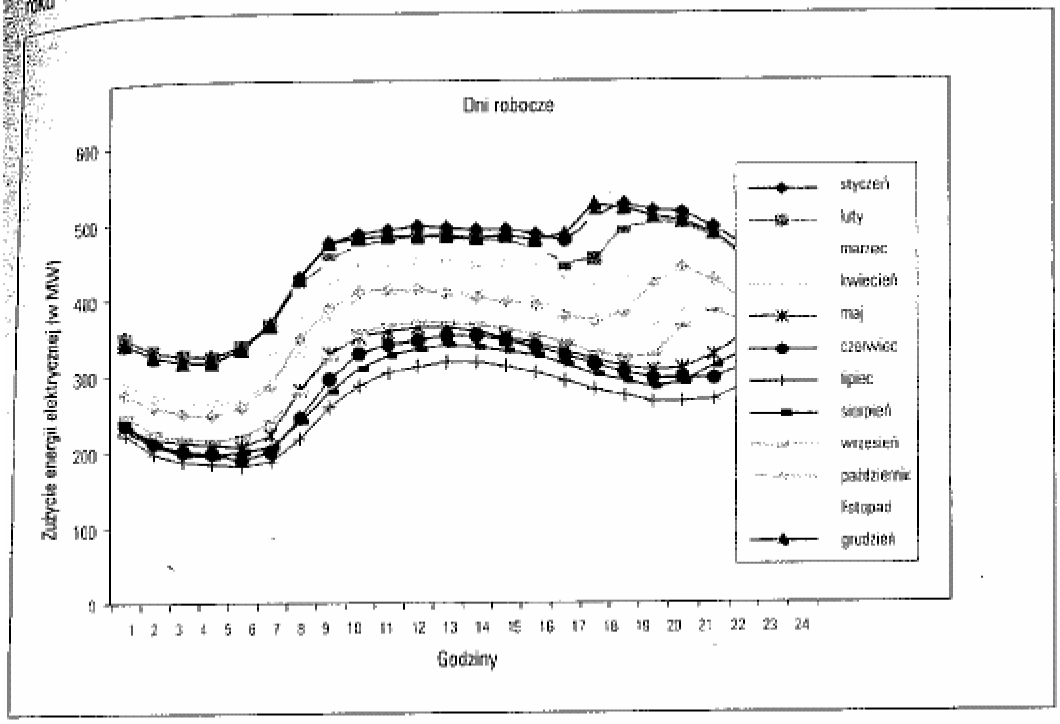
Przykład 6.1

Przeanalizujmy dane[1] dotyczące sprzedaży energii elektrycznej przez Łódzki Zakład Energetyczny w latach 1993–1995. Ponieważ energia elektryczna nie może być magazynowana, wielkość sprzedaży energii jest określona przez zapotrzebowania na to dobro. Na ilustracji 6.1 przedstawiono godzinowe obserwacje dla tej zmiennej.

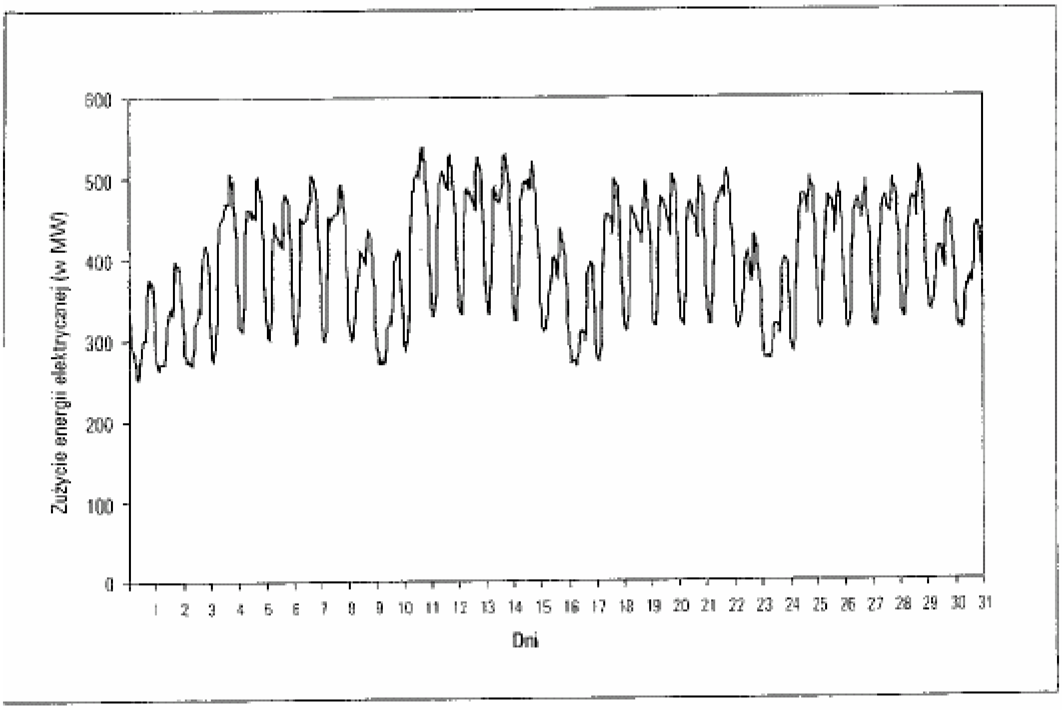
Ilustracja 6.1. Szereg zapotrzebowania na energię elektryczną w Łodzi w latach 1993–1995



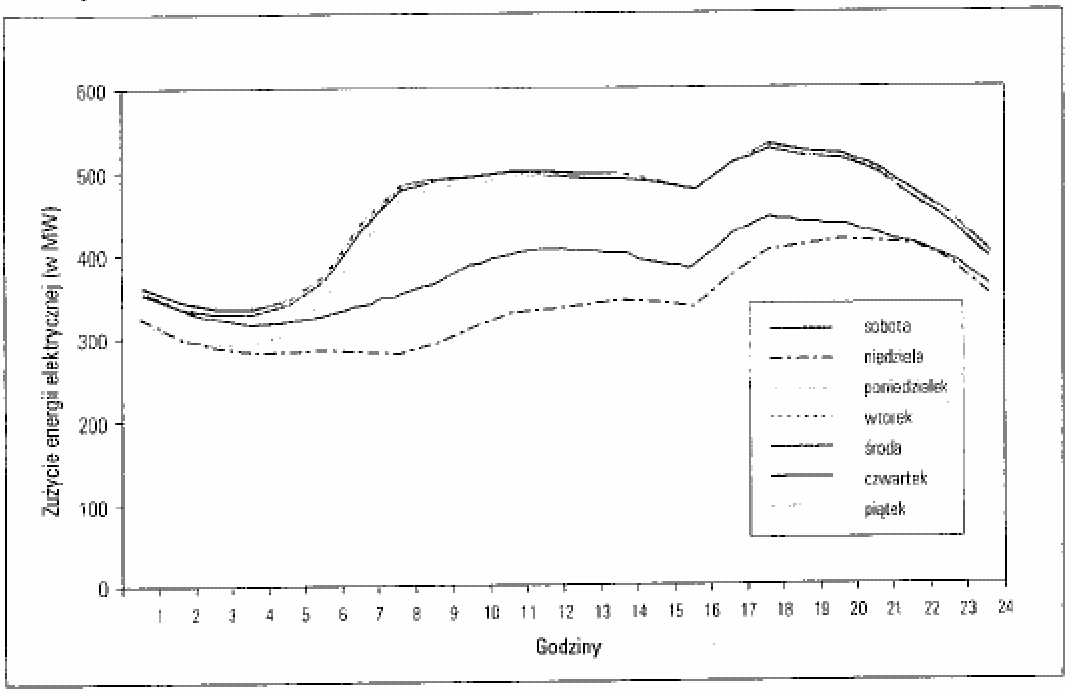
Ilustracja 6.2. Zapotrzebowanie na energię elektryczną w typowym dniu roboczym w kolejnych miesiącach roku



Ilustracja 6.3. Miesięczne zapotrzebowanie na energię elektryczną w styczniu



Ilustracja 6.4. Dobowe zapotrzebowanie na energię elektryczną w kolejnych dniach tygodnia



Założenia odnośnie danych

- Jednostki czasu użyte do pomiarów powinny być równe (szereg okresów).
- Dane historyczne powinny dotyczyć okresu, w którym parametry modelu opisującego proces były (choćby w przybliżeniu) stałe.
- „Im więcej danych, tym lepiej”,
- Problemy z uwzględnianie „kalendarza”, np.
  - standaryzacja długości miesiąca,
  - standaryzacja długości tygodnia,
  - uwzględnianie dni świątecznych.
- Występowanie obserwacji odstających i brakujących:
  - brak pomiaru lub dzień świąteczny,
  - silny wpływ rzadko występującego czynnika (np. jednorazowa realizacja bardzo dużego zamówienia, awaria urządzenia produkcyjnego).

Identyfikacja obserwacji odstających

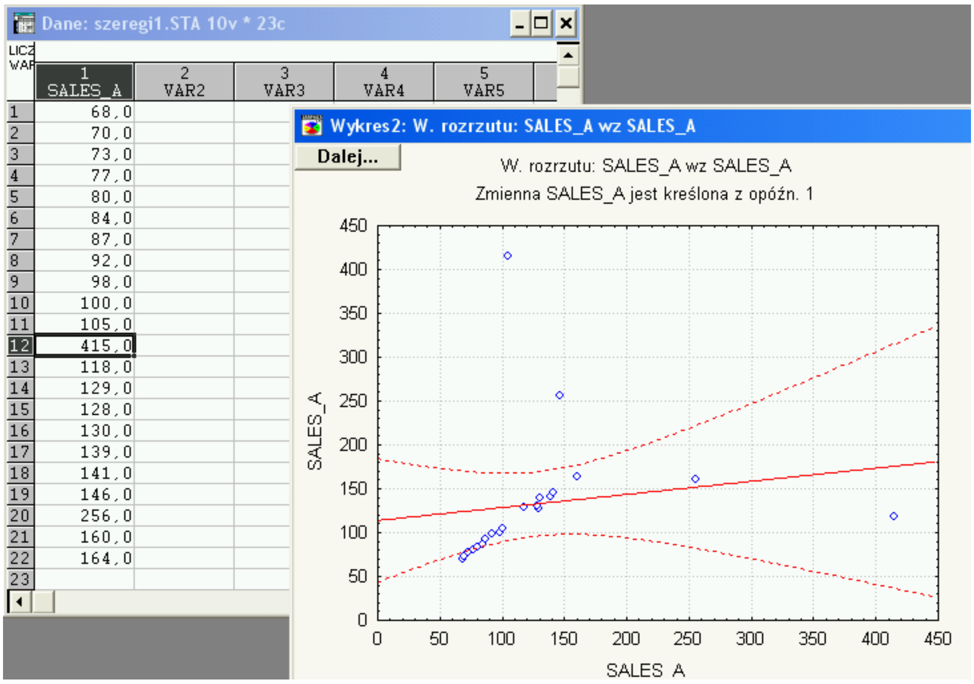
Dla stacjonarnego szeregu czasowego:

Oblicz pierwszy (Q_1) i trzeci (Q_3) kwartyl z szeregu

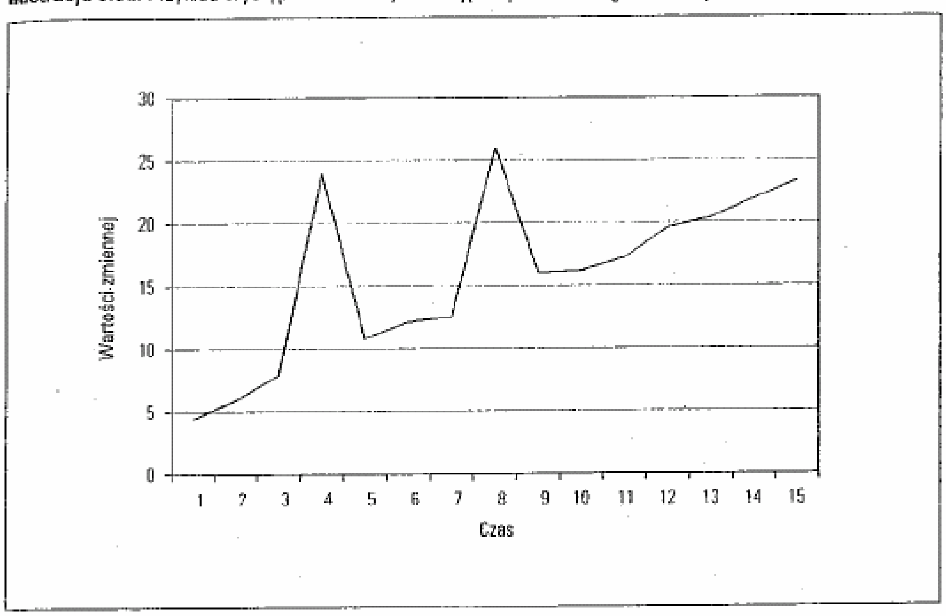
$R_Q = Q_3 - Q_1$; rozstep międzykwartylowy

Obserwacje odstające spoza przedziałów (Q_1-3R_Q; Q_3+3R_Q)

Przykład (Dittmann) sprzedaży miesięcznej produktu A



Ilustracja 6.5a. Przykład występowania danych nietypowych w szeregu czasowym



Podstawowe modele matematyczne

Modelem szeregu czasowego służącym do określenia przyszłej wartości zmiennej prognozowanej *Y* w momencie / okresie *t*, tj. y_t^*, jest model formalny, którego zmiennymi wejściowymi są zmienna czasowa oraz przeszłe wartości lub prognozy zmiennej *Y*.

$$y_t^* = f(t, y_{t-1}, \ldots, y_{t-p}, y_{t-1}^*, \ldots, y_{t-p}^*, \zeta_t)$$

- W trakcie budowy modelu przeprowadza się **dekompozycje szeregu czasowego** w zależności od przyjętych założeń.

W ogólności przyjmuje się addytywną lub multiplikatywną formę modelu.

Model addytywny:

$y_t = f(t) + g(t) + h(t) + \zeta_t$ lub $y_t = const + g(t) + h(t) + \zeta_t$

Model multiplikatywny:

$y_t = f(t) \cdot g(t) \cdot h(t) \cdot \zeta_t$ lub $y_t = const \cdot g(t) \cdot h(t) \cdot \zeta_t$

gdzie *f*(*t*) – funkcja trendu,
g(*t*) – funkcja czasu charakteryzująca wahania sezonowe,
g(*t*) – funkcja czasu charakteryzująca wahania cykliczne,
ζ_t - składnik losowy.

Stosuje się także modele mieszane.

Estymacja trendu

Proste podejścia

- Obserwacja kształtu graficznego wykresu oraz próba doboru funkcji (lub składanego zbioru funkcji)
- Kosztowne, trudne do wykonania; zawodzi dla danych o dużych rozmiarach

Metoda dopasowania (najmniejszych kwadratów)

- Podobna do analizy regresyjnej

Metoda średnich ruchomych

Przykład:

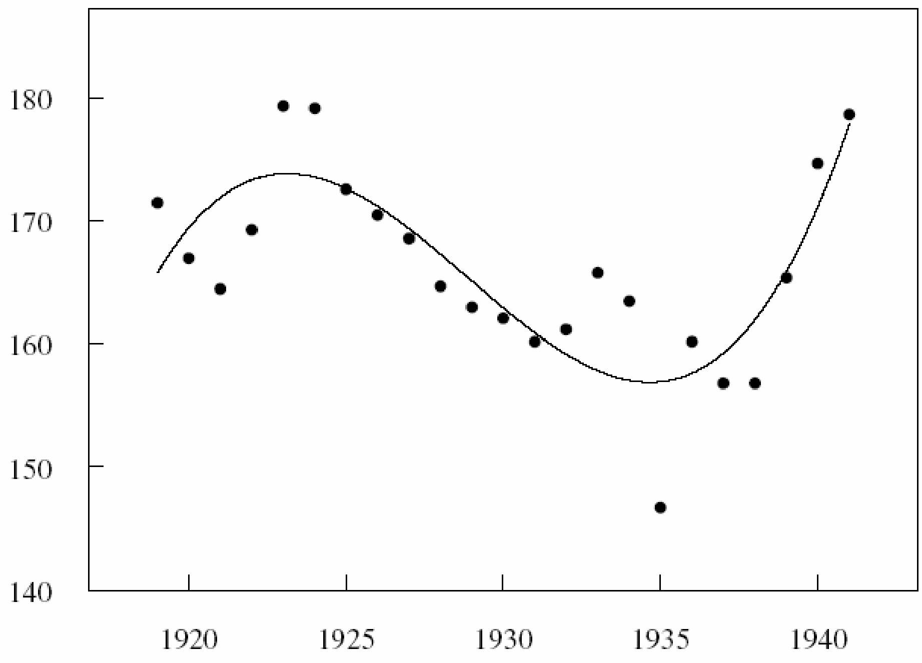


Figure 1. A cubic function fitted to data on meat consumption in the United States, 1919–1941.

Wyodrębnianie i analiza trendu

Dwie grupy metod:

- „mechaniczne” (najczęściej średnie ruchome),
- analityczne (dopasowanie funkcji – aproksymacja; MNK).

Wygładzanie metodami średnich ruchomych

Polega na lokalnym (w czasie) uśrednianiu przebiegu *y*.

Najbardziej popularna **średnia ruchoma**:

- zcentrowana – średnia arytmetyczna *n* pomiarów wokoł punktu *t* (czyli $y_{t\text{ - }n/2},\ldots, y_{t+n/2}$)
- średnia ruchoma **ważona** – preferencja dla aktualnych wartości (większe wagi)

Zamiast średniej arytmetycznej stosuje się także medianę.

Uwaga: dla trendów liniowych do konstrukcji prognozy można także stosować model **podwójnej** średniej ruchomej.

Przykład

średnia ruchoma 3-okresowa $\overline{y}_t = (y_{t-1} + y_t + y_{t+1})/3$

średnia ważona – wagi (1,4,1); np. $\dfrac{1\times 3 + 4\times 7 + 1\times 2}{1+4+1} = 5.5$

Dane oryginalne	3	7	2	0	4	5	9	7	2
Średnia ruchoma		4	3	2	3	6	7	6	
Średnia ważona		5.5	2.5	1	3.5	5.5	8	6.5	

Modele ze stałym poziomem zmiennej prognozowanej

Założenie: poziom wartości zmiennej prognozowanej jest prawie stały w rozpatrywanym okresie (niewielkie odchylenia losowe, brak wahań okresowych).

Idea: wartość zmiennej prognozowanej jest średnią ruchomą z *k* ostatnich wartości tej zmiennej:

$$y_t^* = \frac{1}{k} \sum_{i=t-k}^{t-1} y_i$$

gdzie: *k* – stała wygładzanie.

Problem: jak dobrać *k*? (zalecenia literaturowe; badanie błędu średnio kwadratowego prognozy *ex-post*)

Postarzanie informacji – **średnia ruchoma ważona**

$$y_t^* = \frac{1}{k} \sum_{i=t-k}^{t-1} y_i \cdot w_{i-t+k+1}$$

gdzie $0 < w_1 < \ldots < w_k \leq 1$ oraz $\sum_{i=1}^{k} w_i = 1$

Model **wygładzania wykładniczego**, np.:

$y_t^* = \alpha \cdot y_{t-1} + (1-\alpha) \cdot y_{t-1}^*$, gdzie $\alpha \in (0,1]$.

Kilka uwag o średnich ruchomych (kroczących)

Ogólna forma:

$$\frac{y_1 + y_2 + \cdots + y_n}{n}, \ \frac{y_2 + y_3 + \cdots + y_{n+1}}{n}, \ \frac{y_3 + y_4 + \cdots + y_{n+2}}{n}, \ldots$$

- Wygładza dane (Smoothes the data)
- Może wyeliminować wahania (sezonowe) i nieregularności
- Pomija się część danych (początkowych)
- W pewnym stopniu wrażliwe na obserwacje nietypowe.

Przykład analizy szeregu bez trendu

Dany jest następujący szereg czasowy sprzedanych samochodów w pewnym salonie

Tydzień (t)	Sprzedaż w szt. y_t
1	15
2	17
3	12
4	16
5	15
6	11
7	18
8	17
9	13
10	16

Jakiej sprzedaży samochodów możemy oczekiwać w kolejnych tygodniach, np. w 11 tygodniu?

Średnia arytmetyczna wyznaczona ze wszystkich elementów szeregu – szt.Inne podejście – **metoda średniej ruchomej**
Średnia czteroelementowa:

Średnia sprzedaż tygodniowa (tygodnia 1-4)$= \dfrac{15+17+12+16}{4} = 15$ szt.

Jest to równocześnie prognoza sprzedaży dla piątego tygodnia.

Przykład – oszacowywanie błędu prognozy

Błąd prognozy dla chwili *t* jest równy $\delta_t = y_t - y_t^*$

Syntetyczna charakterystyka dokładności prognoz, np. błąd średniokwadratowy *MSE*

$$MSE = \frac{\sum_{t=1}^{n} \delta_t}{n}$$

Obliczmy *MSE* dla prognozy uzyskanej czteroelementową średnią ruchomą:

oraz dla pięcioelementowej średniej ruchomej:

Przykład sprzedaży samochodów
– szereg czasowy bez trendu.

Średnia ruchoma i błędy prognozy

t	Y_t	Średnia ruchoma cztero-elementowa	Błąd prognozy	(Błąd)^2	Średnia ruchoma pięcio-elementowa	Błąd prognozy	(Błąd)^2
1	2	3	4	5	6	7	8
1	15	–	–	–	–	–	–
2	17	–	–	–	–	–	–
3	12	–	–	–	–	–	–
4	16	–	–	–	–	–	–
5	15	15,0	0	0	0	–	–
6	11	15,0	–4,0	16	15,0	–4	16,00
7	18	13,5	4,5	20,25	14,2	3,8	14,44
8	17	15,0	2,0	4,0	14,4	2,6	6,67
9	13	15,2	–2,2	4,84	15,4	–2,4	5,76
10	16	14,75	1,25	1,563	14,8	1,2	1,44
11	15	16		Σ=46,653	15		Σ=44,4

Źródło: Na podstawie danych z przykładu 5.1.

Wyrównywanie wykładnicze Browna

$$y_t^* = \alpha \cdot y_{t-1} + (1-\alpha) \cdot y_{t-1}^* \ ,$$

lub w przekształconej postaci $y_t^* = y_{t-1}^* + \alpha \cdot (y_{t-1} - y_{t-1}^*)$

Rozważmy dwie różne wartości α=0,2 oraz 0,8.

Tabela 5.3

Wyniki metody wyrównywania wykładniczego dla danych z przykładu 5.1

t	Y_t	F_t $\alpha = 0,2$	$Y_t - F_t$ $\alpha = 0,2$	$(Y_t - F_t)^2$ $\alpha = 0,2$	F_t $\alpha = 0,8$	$Y_t - F_t$ $\alpha = 0,8$	$(Y_t - F_t)^2$ $\alpha = 0,8$
1	2	3	4	5	6	7	8
1	15	15,0	0,0	0	15,0	0	0
2	17	15,0	2,0	4,0	15,0	2,0	4,0
3	12	15,4	–3,4	11,56	16,6	–4,6	21,16
4	16	14,7	1,3	1,69	12,9	3,1	9,61
5	15	15,0	0,0	0	15,4	–0,4	0,16
6	11	15,0	–4,0	16,0	15,1	–4,1	16,81
7	18	14,2	3,8	14,44	11,8	6,2	38,44
8	17	15,0	2,0	4,0	16,0	0,2	0,04
9	13	15,4	–2,4	5,76	17,0	–4,0	16,0
10	16	15,0	1,0	1,0	13,8	2,2	4,84
11		15,2		58,45	15,6		111,04

Źródło: Obliczenia własne na podstawie danych z przykładu 5.1.

Modele analityczne szeregów czasowych ze trendem

Wybór modelu: np. addytywny $y_t = f(t) + \zeta_t$ lub multiplikatywny $y_t = f(t) \cdot \zeta_t$

Znalezienie funkcji *f*(*t*) najlepiej pasującej do wyrazów szeregu zmiennej prognozowanej – przyjęcie hipotezy, co do określonej postaci funkcji trendu.

Często stosowane funkcja liniowa: $y_t = \alpha + \beta \cdot t$.

Ponadto dla prognoz krótkoterminowych:

- funkcja wykładnicza: $y_t = e^{\alpha + \beta \cdot t}$ lub $y_t = \alpha \cdot \beta^t$
- wielomiany stopnia drugiego: $y_t = \alpha_0 + \alpha_1 \cdot t + \alpha_2 \cdot t^2$ (α2>0)
- potęgowa: $y_t = \alpha \cdot t^{\beta}$ (β>1).

Funkcje o malejących przyrostach, np.:

- funkcja logarytmiczna: $y_t = \alpha + \beta \cdot \ln t$ (β>1).
- wielomian odwrotnościowy: $y_t = \alpha_0 + \alpha_1 \cdot t^{-1} + \alpha_2 \cdot t^{-2}$ (α2<0).
- funkcja liniowo-odwrotnościowa: $y_t = \alpha + \dfrac{\beta}{t}$ (β<0).
- funkcja ilorazowa: $y_t = \dfrac{\alpha \cdot t}{\beta + t}$ (α,β>0).
- funkcja logistyczna: $y_t = \dfrac{\alpha}{1 + \beta \cdot e^{-\delta \cdot t}}$

Regresja liniowa

Funkcja w postaci: $f(t) = \alpha + \beta \cdot t$

Wartości parametrów α oraz β poszukujemy za pomocą minimalizacji składnika resztowego (MNK):
$\min \sum_t (y_t - f(t))^2$

Ocena jakości dopasowania

Modele liniowe:

S = odchylenie standardowe składnika resztowego – przeciętne odchylenie zaobserwowanej wartości rzeczywistej y_t od odpowiadającym im wartościom teoretycznym $\hat{y}_t$ wyznaczonym z modelu.

w = współczynnik zmienności losowej (wyrazistości)

R^2 = współczynnik determinacji

Testowanie hipotez o istotności współczynników (F)

Dla modeli nieliniowych – powyższe techniki oceny powinny być obliczane i interpretowane tylko dla postaci transformowanej zlinearyzowanej.

Ponadto można badać rozkład odchyleń losowych modelu.

Przykład analizy szeregu czasowego z trendem

Analiza sprzedaży obuwia produkowanego przez pewne przedsiębiorstwo.

Tablica 5.3
Sprzedaż obuwia dziecięcego przedsiębiorstwa A
w latach 1991–1998

Lata	Sprzedaż w mln zł
1991	3
1992	4
1993	6
1994	10
1995	11
1996	12
1997	14
1998	18

Źródło: Dane umowne.

Zastosujmy model liniowy

$$y_t^* = b_0 + b_1 \cdot t$$

Obliczmy b_0 i b_1 na podstawie wzorów wynikających z metody MNK

$$y_t^* = 1,51 + 1,83 \cdot t$$

Przykład prognozy na 1999 rok

$$y_9 = 1,51 + 1,83 \cdot 9 = 17,98 \text{ mln zł.}$$

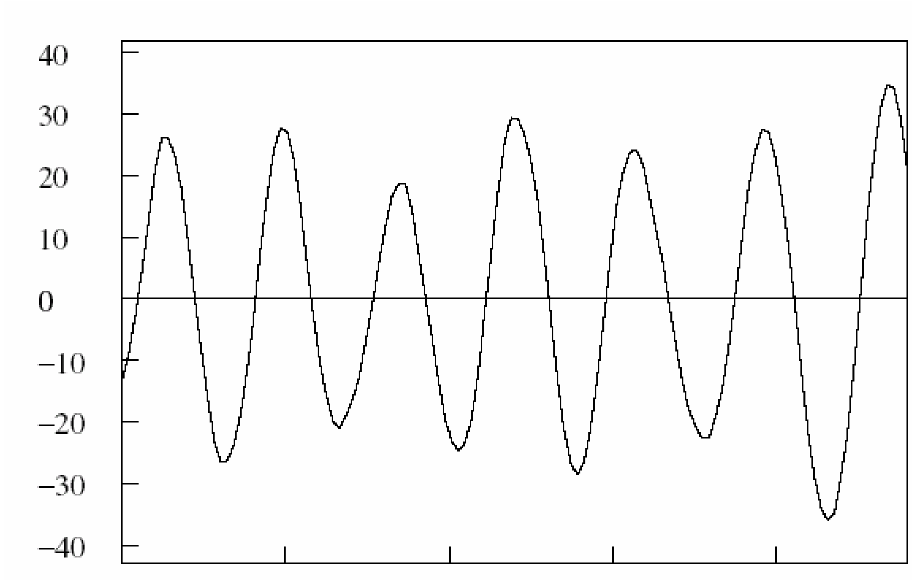
Analiza okresowości

W przebiegu szeregu czasowego dostrzega się pewne wahania, powtarzające się w tych samych mniej więcej rozmiarach (bezwzględnych lub względnych), co jakiś w przybliżeniu stały odstęp czasu.

Cel: wyodrębnienie składowej okresowej (sezonowej, cyklicznej).

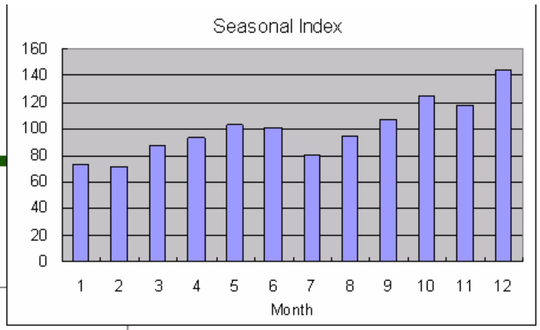
Różne metody:

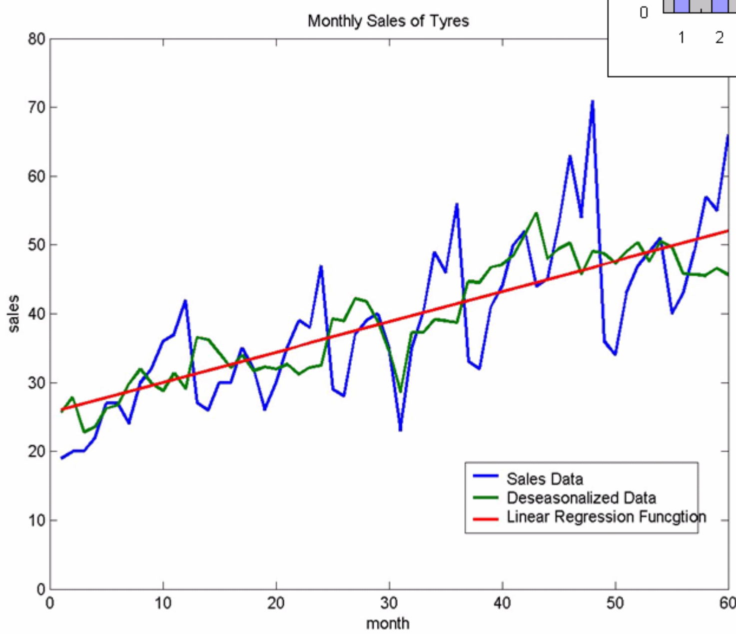
- okres zmian znany – klasyczna dekompozycja szeregu czasowego (np. metoda wskaźników, analiza harmoniczna, modele wygładzania wykładniczego np. Wintersa),
- okres zmian nieznany – modele z opóźnionymi wartościami zmiennej prognozowanej (np. modele ARIMA, analiza autokorelacji).



Przykład sezonowości w danych handlowych

Seasonal Index





Raw data from http://www.bbk.ac.uk/manop/man/docs/QII_2_2003%20Time%20series.pdf

10

- Seasonal index
  - Set of numbers showing the relative values of a variable during the months of the year
  - E.g., if the sales during October, November, and December are 80%, 120%, and 140% of the average monthly sales for the whole year, respectively, then 80, 120, and 140 are seasonal index numbers for these months
- Deseasonalized data
  - Data adjusted for seasonal variations for better trend and cyclic analysis
  - Divide the original monthly data by the seasonal index numbers for the corresponding months

Modele szeregów czasowych z wahaniami sezonowymi

Wahania okresowe tworzą cykl **sezonowy**
– długość cyklu / okres wahań.

Okres / sezon składa się z **faz** (kształtowanie się przebiegu, np. szybki wzrost, lekki wzrost, spadek, itd.)

Liczba faz w cyklu, np. 12 faz dotyczących danych miesięcznych, 4 faz dotyczących danych kwartalnych -> długość okresu (sezonu).

Metoda wskaźników:

Polega na wyznaczeniu **wskaźników sezonowości** poszczególnych faz cyklu.

Amplitudy wahań (wynikające z porównania wartości rzeczywistych zmiennej z wartościami teoretycznymi uzyskanymi z modelu) dzieli się na:

1. Bezwzględnie stałe (w analogicznych fazach cyklu).
2. Względnie stałe (wielkości amplitud zmieniają się w przybliżeniu w tym samym stosunku).

Okresowość może mieć względem trendu charakter **addytywny** (w przypadku 1 bezwzględnych stałych wahań) lub **multiplikatywny** (w przypadku 2).

Metoda wskaźników, cz. 2

Model addytywny: $y_{ti} = \hat{y}_{ti} + s_i + \zeta_t$

Model multiplikatywny: $y_{ti} = \hat{y}_{ti} \cdot s_i \cdot \zeta_t$

gdzie: $t = 1,2,\ldots,n$; $i=1,2, \ldots,r$,

y_{ti} – rzeczywista wartość zmiennej w okresie t w i-tej fazie cyklu,

$\hat{y}_{ti}$ – teoretyczna wartość zmiennej w okresie t w i-tej fazie cyklu,

s_i - wskaźnik sezonowości dla i-tej fazie cyklu,

ζ - składnik losowy,

r – liczba faz cyklu.

Sposób postępowania:

- wyodrębnienie trendu,
- eliminacja trendu z szeregu czasowego,
- eliminacja wahań przypadkowych,
- obliczenie wskaźników sezonowości

Eliminacja trendu: $z_{ti} = y_{ti} - \hat{y}_{ti}$ (addytywny) lub $z_{ti} = \dfrac{y_{ti}}{\hat{y}_{ti}}$

Wskaźniki okresowości

Eliminacja oddziaływania składnika losowego $\rightarrow$ obliczenie surowych wskaźników sezonowości.

$$z_i = \frac{1}{k} \sum_{j=0}^{k-1} z_{i+j \cdot r, i}$$

gdzie k – liczba jednoimiennych faz w badanym szeregu czasowym, i – wybrana faza, r – liczba faz w okresie.

Wskaźniki podlegając korekcji:

$s_i = z_i - q$ (dla modelu addytywnego)

lub $s_i = \frac{z_i}{q}$ (dla modelu multiplikatywnego)

gdzie $q = \frac{1}{r} \sum_{i=1}^{r} z_i$.

Suma wskaźników sezonowości powinna być równa zeru (model addytywny) lub liczbie faz tworzących okres (model multiplikatywny).

Przykład wyznaczania wskaźników sezonowości

Tabela 4.1

Dzienna sprzedaż mleka (w litrach)

Tydzień	Dzień tygodnia				
	poniedziałek	wtorek	środa	czwartek	piątek
Pierwszy	63	47	43	45	30
Drugi	65	48	44	46	31
Trzeci	68	50	48	50	33

Źródło: Raporty sklepowe.

Analiza dziennej sprzedaży mleka (Dittmann, 1998)
Zauważ, wahania sezonowe – okres tygodnia
Zastosujmy model multiplikatywny i dwie metody

Metoda I – wygładzenie średnią ruchomą (niecentrowaną, 5-elementową)

Tabela 4.2

Wygładzanie szeregu niecentrowaną średnią ruchomą

Tydzień	Dzień tygodnia	Obserwacje	Niecentrowana średnia ruchoma	Wygładzenie
Pierwszy	poniedziałek	63	–	–
	wtorek	47	–	–
	środa	43	45,6	0,943
	czwartek	45	46,0	0,978
	piątek	30	46,2	0,649
Drugi	poniedziałek	65	46,4	1,401
	wtorek	48	46,6	1,030
	środa	44	46,8	0,940
	czwartek	46	47,4	0,970
	piątek	31	47,8	0,649
Trzeci	poniedziałek	68	48,6	1,399
	wtorek	50	49,4	1,012
	środa	48	49,8	0,964
	czwartek	50	–	–
	piątek	33	–	–

Źródło: Obliczenia własne.

Tabela 4.3

Surowe wskaźniki sezonowości

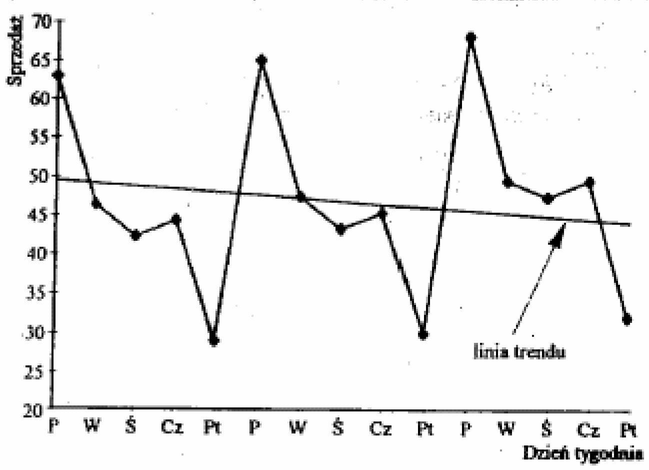
Tydzień	Dzień tygodnia				
	poniedziałek	wtorek	środa	czwartek	piątek
Pierwszy	–	–	0,943	0,978	0,649
Drugi	1,401	1,030	0,940	0,970	0,649
Trzeci	1,399	1,012	0,964	–	–
Średnie	1,400	1,021	0,949	0,974	0,649

Źródło: Obliczenia własne.

q – współczynnik korygujący $\dfrac{1,4+1,021+0,949+0,974+0,649}{5} = 0,9986$

wskaźniki sezonowości:

Dzień	Poniedziałek	Wtorek	Środa	Czwartek	Piątek
Wskaźnik	1,402	1,023	0,950	0,975	0,650



Rys. 4.4. Dzienna sprzedaż mleka (w litrach)

MNK – funkcja trendu

$\hat{y}_t = 50,3 - 0,368 \cdot t$

II Metoda analityczna

Tabela 4.4

Wygładzanie szeregu funkcją trendu

Tydzień	Dzień tygodnia	Obserwacje	Wartości trendu	Wygładzenie
Pierwszy	poniedziałek	63	49,9	1,263
	wtorek	47	49,6	0,948
	środa	43	49,2	0,874
	czwartek	45	48,8	0,922
	piątek	30	48,5	0,619
Drugi	poniedziałek	65	48,1	1,351
	wtorek	48	47,7	1,006
	środa	44	47,4	0,928
	czwartek	46	47,0	0,979
	piątek	31	46,6	0,665
Trzeci	poniedziałek	68	46,3	1,469
	wtorek	50	45,9	1,089
	środa	48	45,5	1,055
	czwartek	50	45,1	1,109
	piątek	33	44,8	0,737

Źródło: Obliczenia własne.

Tabela 4.5

Surowe wskaźniki sezonowości

Tydzień	Dzień tygodnia				
	poniedziałek	wtorek	środa	czwartek	piątek
Pierwszy	1,263	0,948	0,874	0,922	0,619
Drugi	1,351	1,006	0,928	0,979	0,655
Trzeci	1,469	1,089	1,055	1,109	0,737
Średnie	1,361	1,014	0,952	1,003	0,674

Źródło: Obliczenia własne.

$q = 0,9992$

wskaźniki sezonowości:

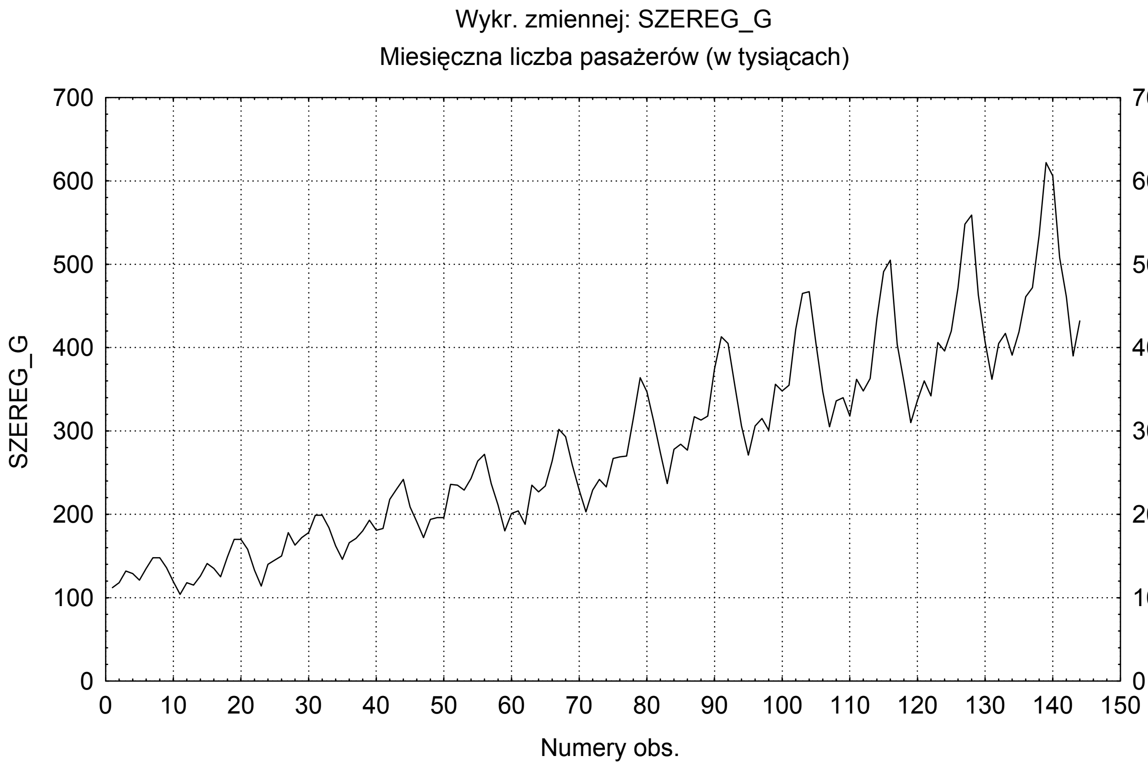
Dzień	Poniedziałek	Wtorek	Środa	Czwartek	Piątek
Wskaźnik	1,360	1,013	0,952	1,002	0,673

Przykład analizy szeregu „G”

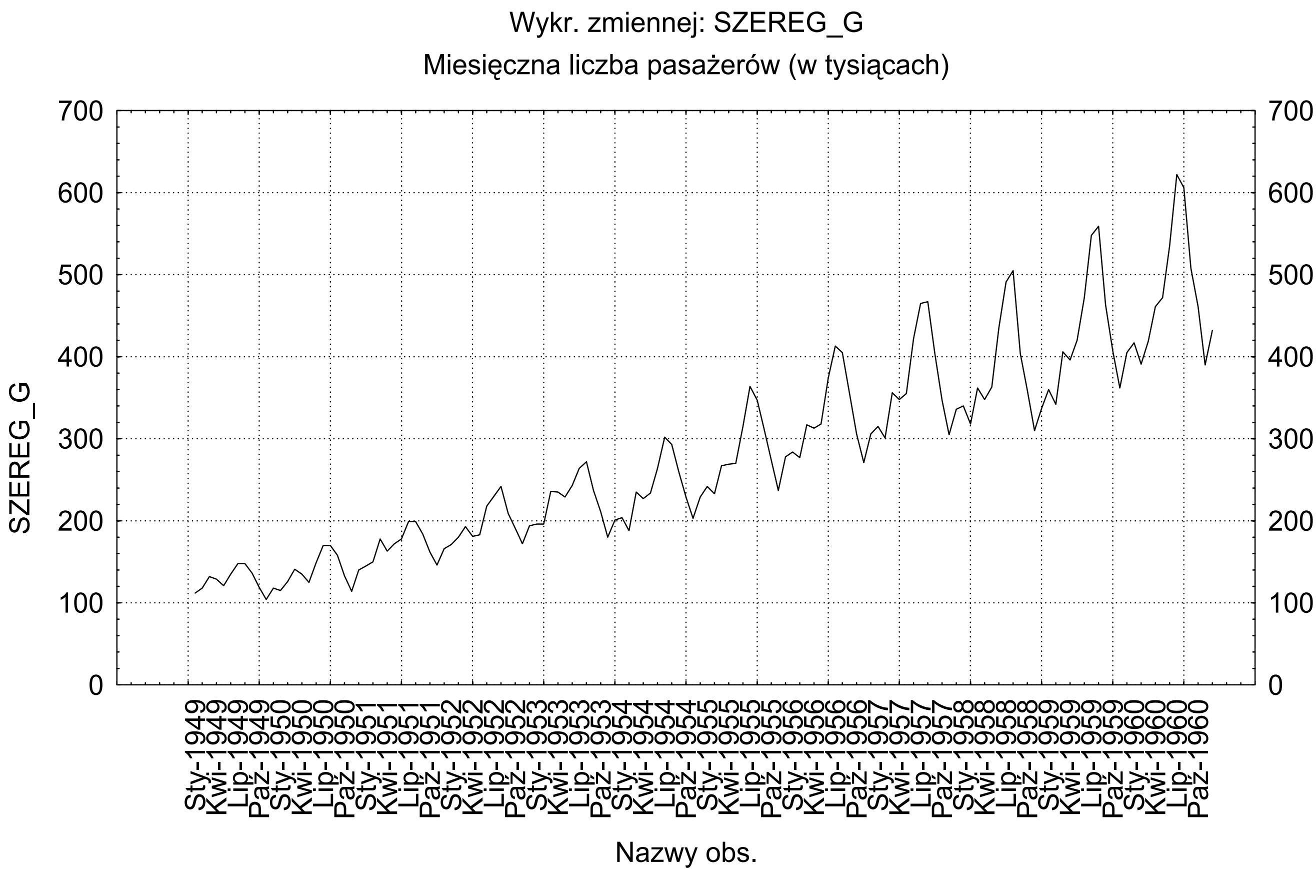
Podany w (Box i Jenkins, 1976, str. 531) reprezentuje miesięczne liczby (mierzone w tysiącach) pasażerów międzynarodowej linii lotniczej w kolejnych dwunastu latach od 1949 do 1960.

	SZEREG_G
Sty-1949	112000
Lut-1949	118000
Mar-1949	132000
Kwi-1949	129000
Maj-1949	121000
Cze-1949	135000
Lip-1949	148000
Sie-1949	148000
Wrz-1949	136000
Paź-1949	119000
Lis-1949	104000
Gru-1949	118000
Sty-1950	115000
Lut-1950	126000
…	…
Sie-1960	606000
Wrz-1960	508000
Paź-1960	461000
Lis-1960	390000
Gru-1960	432000

Szereg G



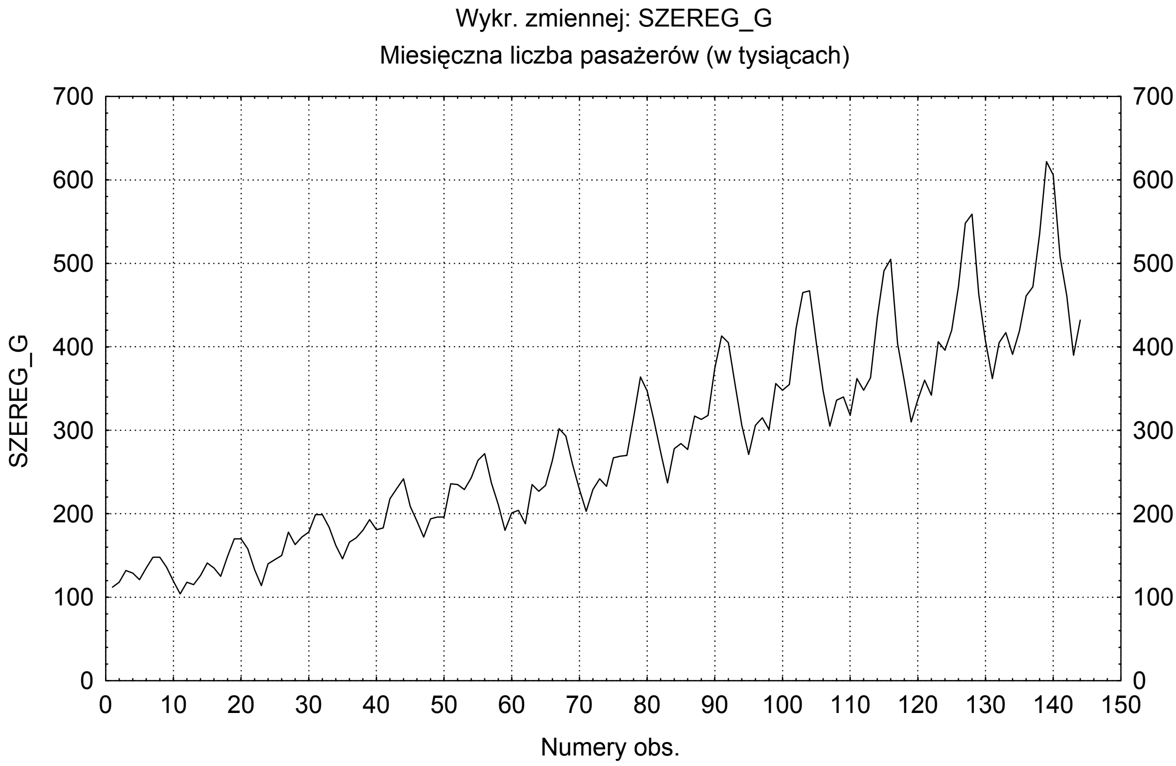
Wykr. zmiennej: SZEREG_G
Miesięczna liczba pasażerów (w tysiącach)
SZEREG_G
Numery obs.



Wykr. zmiennej: SZEREG_G
Miesięczna liczba pasażerów (w tysiącach)
SZEREG_G
Nazwy obs.

Seria G

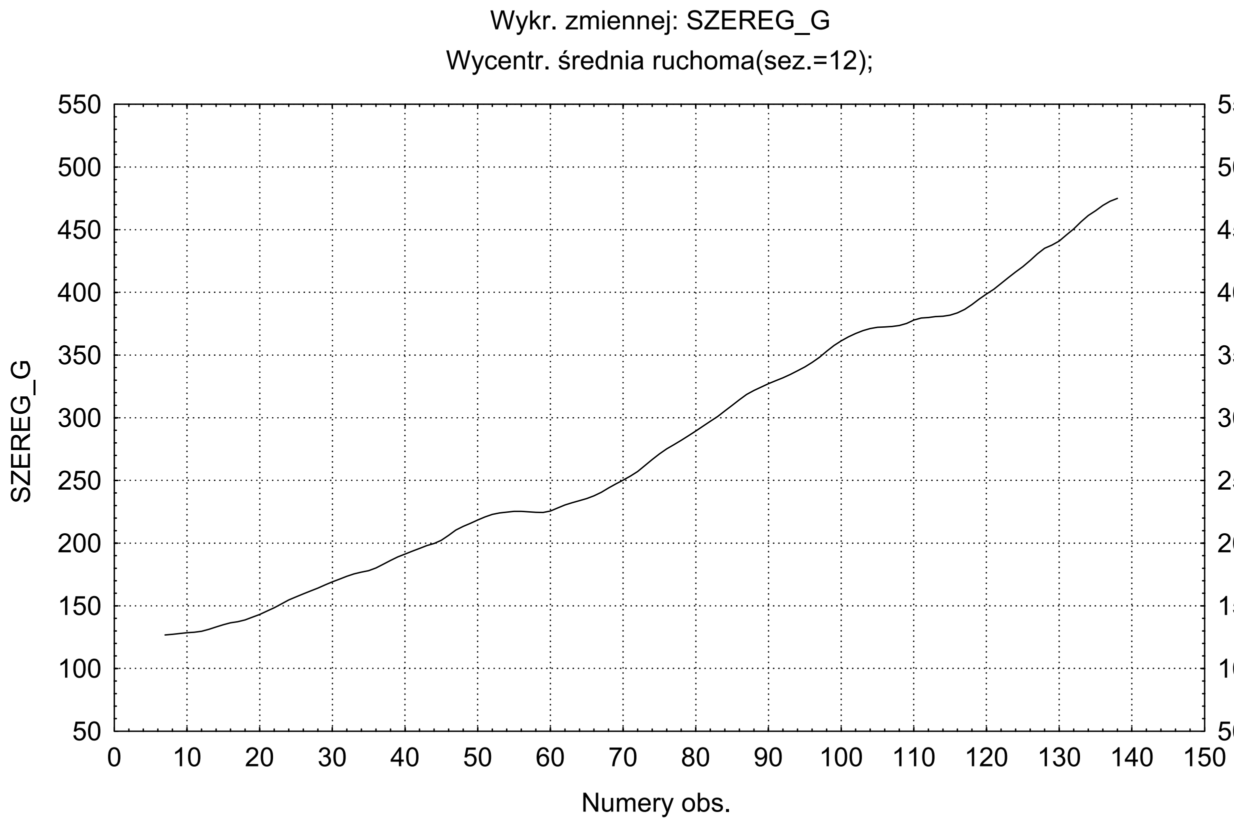
Na wykresie danych rocznych liczby pasażerów linii lotniczej widać prawie liniowy trend wskazujący, że linia lotnicza cieszyła się równomiernym wzrostem pasażerów w badanym okresie.



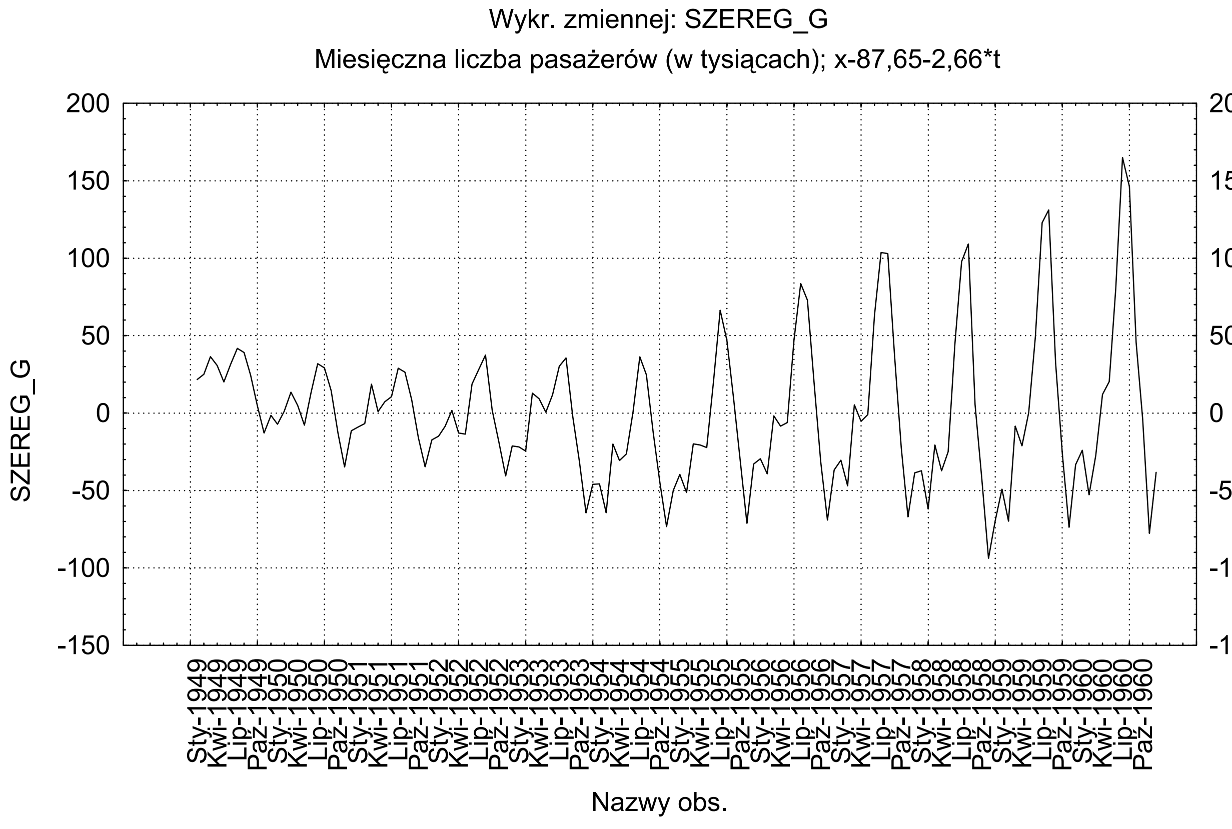
Dane miesięczne odpowiadają co roku prawie identycznemu wzorcowi (np. więcej osób podróżuje podczas wakacji niż w innych porach roku).

Amplituda zmian sezonowych wzrasta wraz z ogólnym trendem (tzn. wariancja jest skorelowana ze średnią segmentów szeregu). Tego typu sezonowość nazywana jest sezonowością multiplikatywną. Względna amplituda zmian sezonowych jest tu stała w czasie, to znaczy, że wahania sezonowe są proporcjonalne do trendu.

Przykład użycia modelu wahań sezonowych dla szeregu G



Model analityczny = 87,65-2,66·t



Przykład dekompozycji sezonowej (metoda analityczna)

D.Witkowska (str. 154) dane nt. kwartalnych nakładów inwestyczynych w latach 1992-1997

Trend liniowy: $\hat{y}_t = 1880{,}95 + 75{,}28 \cdot t$

Sezonowość: kwartalna (r = 4); 6 kolejnych sezonów

Surowe wskaźniki:

Z_1=0,661; Z_2=0,9; Z_1=0,939; Z_1=1,468;

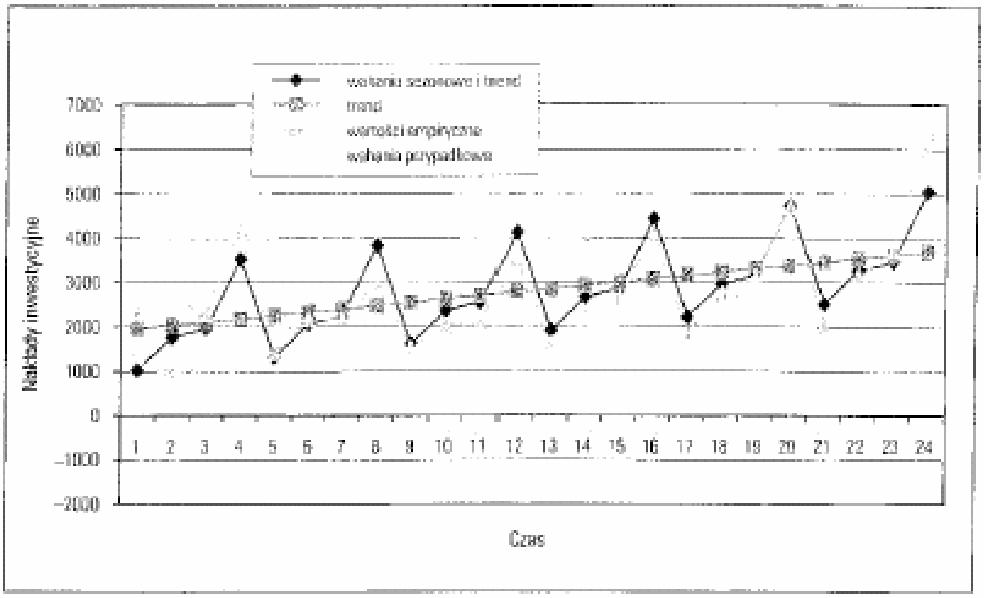
Skorygowane wskaźniki sezonowości:

S_1=0,667; S_2=0,907; Z_1=0,946; Z_1=1,48;

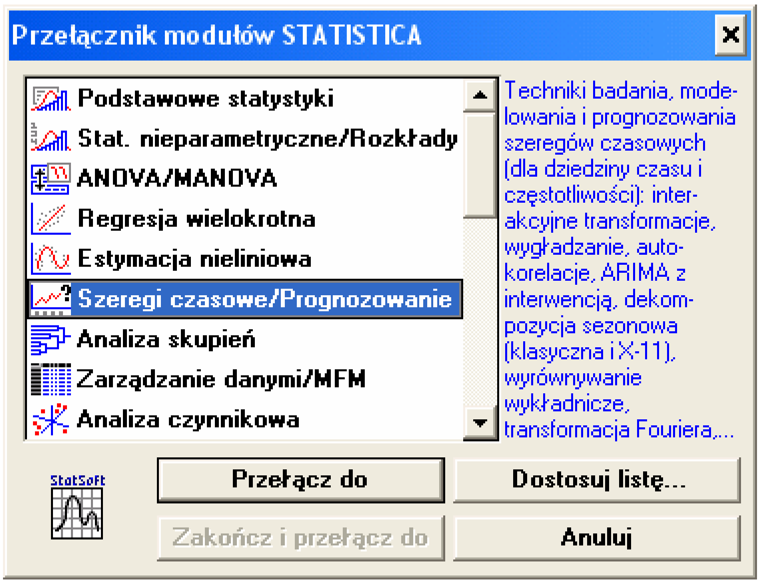
Tabela 6.2. Obliczenia do przykładu

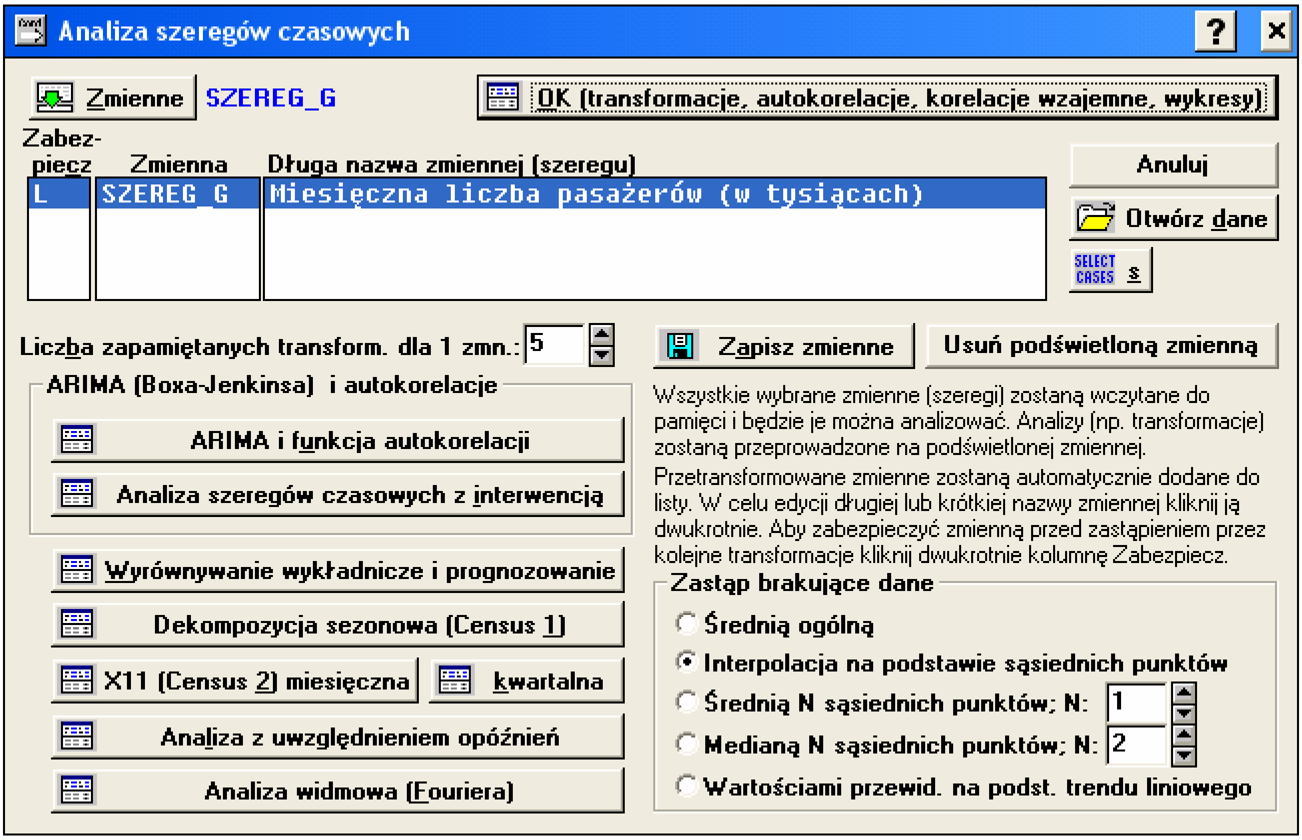
Okresy	t	y_t	$\hat{y}_t$	S_i^*
1	2	3	4	5
1992, I kwartał	1	2270,50	1956,23	0,67
1992, II kwartał	2	2656,83	2031,50	0,91
1992, III kwartał	3	2285,21	2106,78	0,95
1992, IV kwartał	4	4206,14	2182,05	1,48
1993, I kwartał	5	1424,96	2257,33	0,67
1993, II kwartał	6	2165,40	2332,60	0,91
1993, III kwartał	7	2193,31	2407,88	0,95
1993, IV kwartał	8	2976,59	2483,15	1,48
1994, I kwartał	9	1545,18	2558,43	0,67
1994, II kwartał	10	2032,53	2633,71	0,91
1994, III kwartał	11	2117,94	2708,98	0,95
1994, IV kwartał	12	3488,68	2784,26	1,48
1995, I kwartał	13	1618,58	2859,53	0,67
1995, II kwartał	14	2248,79	2934,81	0,91
1995, III kwartał	15	2672,11	3010,08	0,95
1995, IV kwartał	16	4126,51	3085,36	1,48
1996, I kwartał	17	1847,15	3160,63	0,67
1996, II kwartał	18	2754,43	3235,91	0,91
1996, III kwartał	19	3125,50	3311,19	0,95
1996, IV kwartał	20	4719,98	3386,46	1,48
1997, I kwartał	21	2040,60	3461,74	0,67
1997, II kwartał	22	3174,23	3537,01	0,91
1997, III kwartał	23	3710,47	3612,29	0,95
1997, IV kwartał	24	6323,84	3687,56	1,48
Suma	×	67725,46	×	×

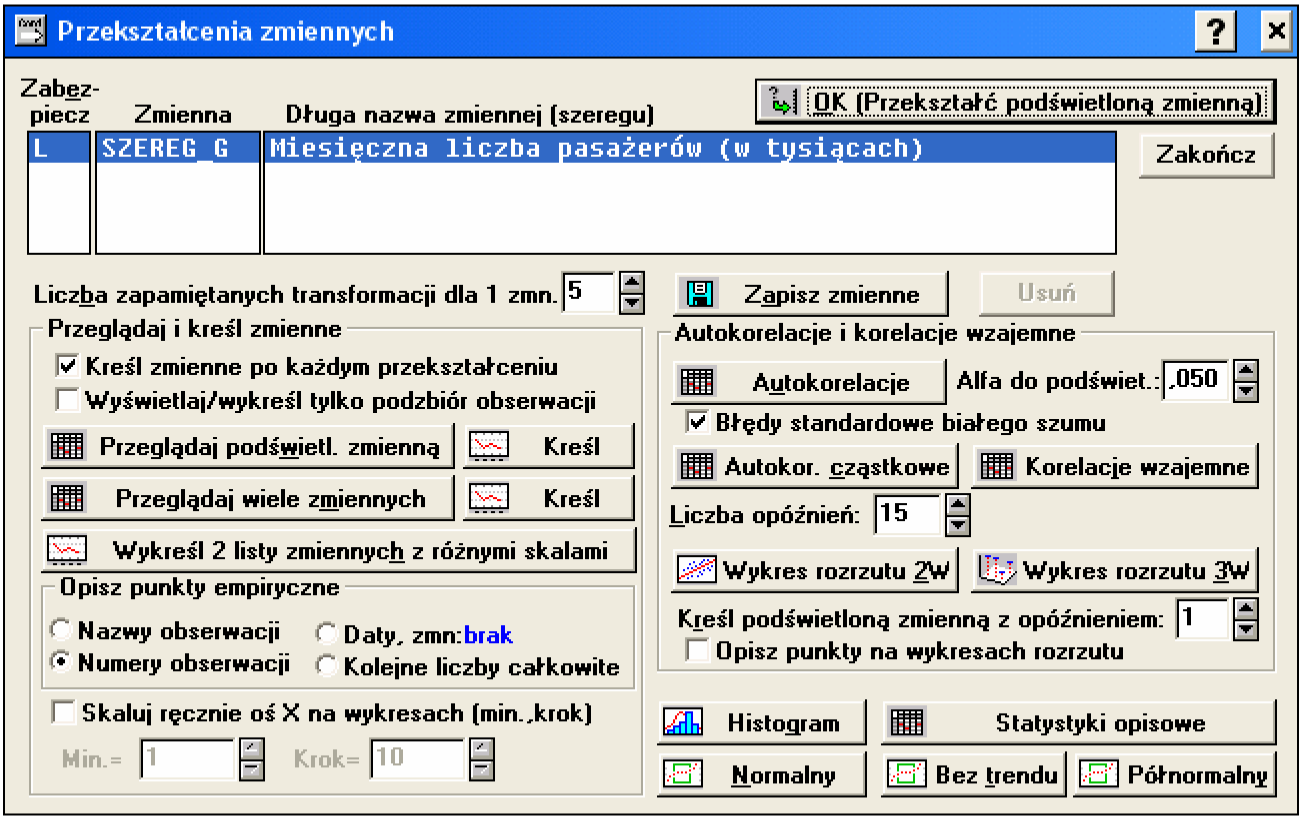
Ilustracja 6.7. Dekompozycja szeregu czasowego za pomocą metody analitycznej

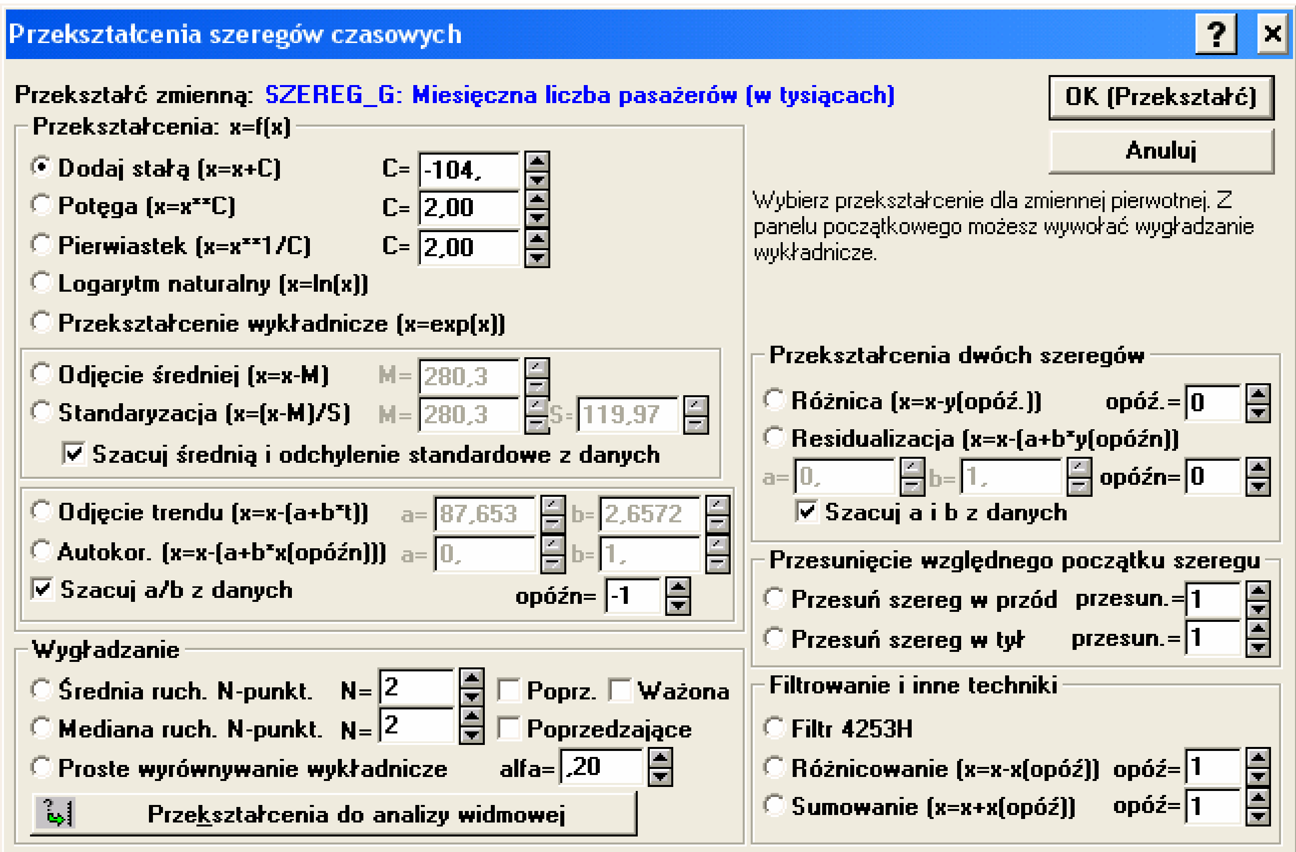


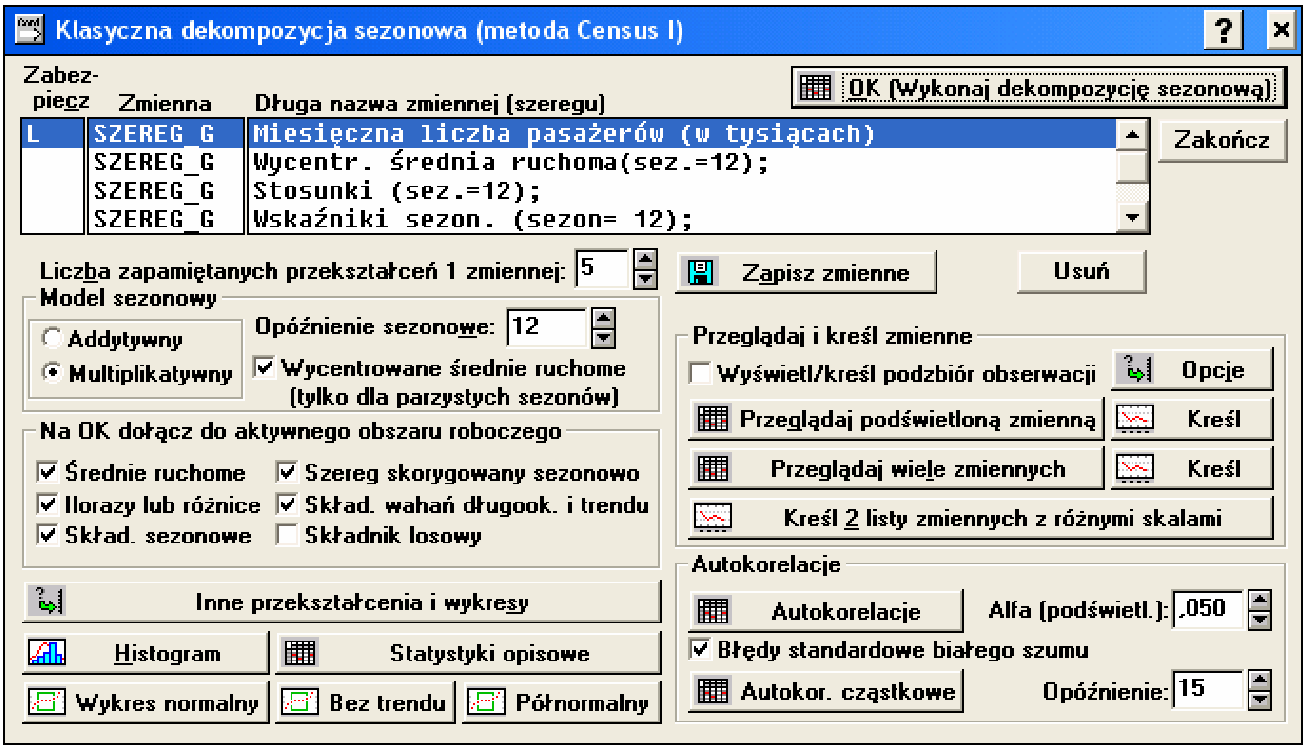
Statsoft Statistica – gdzie szukać?

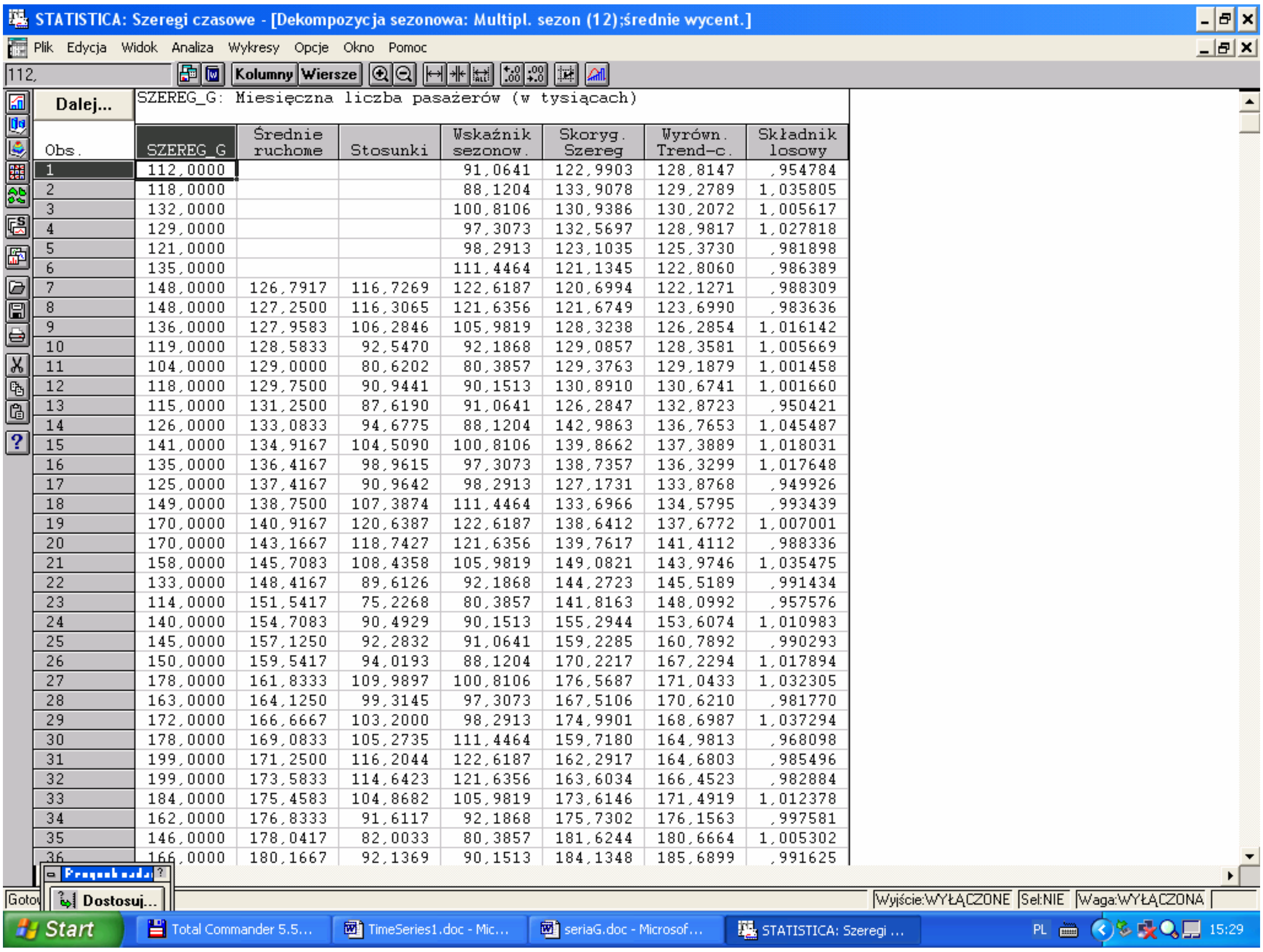





Przekształcenia zmiennych
Zabez-piecz
Zmienna
Długa nazwa zmiennej (szeregu)
OK (Przekształć podświetloną zmienną)
L
SZEREG_G
Miesięczna liczba pasażerów (w tysiącach)
Zakończ
Liczba zapamiętanych transformacji dla 1 zmn.
5
Zapisz zmienne
Usuń
Przeglądaj i kreśl zmienne
Kreśl zmienne po każdym przekształceniu
Wyświetlaj/wykreśl tylko podzbiór obserwacji
Przeglądaj podświetl. zmienną
Kreśl
Przeglądaj wiele zmiennych
Kreśl
Wykreśl 2 listy zmiennych z różnymi skalami
Opisz punkty empiryczne
Nazwy obserwacji
Daty, zmn:brak
Numery obserwacji
Kolejne liczby całkowite
Skaluj ręcznie oś X na wykresach (min.,krok)
Min.=
1
Krok=
10
Autokorelacje i korelacje wzajemne
Autokorelacje
Alfa do podświet.:
,050
Błędy standardowe białego szumu
Autokor. cząstkowe
Korelacje wzajemne
Liczba opóźnień:
15
Wykres rozrzutu 2W
Wykres rozrzutu 3W
Kreśl podświetloną zmienną z opóźnieniem:
1
Opisz punkty na wykresach rozrzutu
Histogram
Statystyki opisowe
Normalny
Bez trendu
Półnormalny


Przekształcenia szeregów czasowych
Przekształć zmienną: SZEREG_G: Miesięczna liczba pasażerów (w tysiącach)
OK (Przekształć)
Anuluj
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(opóźn)))
a=
0,
b=
1,
Szacuj a/b z danych
opóźn=
-1
Wygładzanie
Średnia ruch. N-punkt.
N=
2
Poprz.
Ważona
Mediana ruch. N-punkt.
N=
2
Poprzedzające
Proste wyrównywanie wykładnicze
alfa=
,20
Przekształcenia do analizy widmowej
Wybierz przekształcenie dla zmiennej pierwotnej. Z panelu początkowego możesz wywołać wygładzanie wykładnicze.
Przekształcenia dwóch szeregów
Różnica (x=x-y(opóź.))
opóź.=
0
Residualizacja (x=x-(a+b*y(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(opóź))
opóź=
1
Sumowanie (x=x+x(opóź))
opóź=
1





Obs.	SZEREG_G	Średnie ruchome	Stosunki	Wskaźnik sezonow.	Skoryg. Szereg	Wyrówn. Trend-c.	Składnik losowy
1	112,0000			91,0641	122,9903	128,8147	,954784
2	118,0000			88,1204	133,9078	129,2789	1,035805
3	132,0000			100,8106	130,9386	130,2072	1,005617
4	129,0000			97,3073	132,5697	128,9817	1,027818
5	121,0000			98,2913	123,1035	125,3730	,981898
6	135,0000			111,4464	121,1345	122,8060	,986389
7	148,0000	126,7917	116,7269	122,6187	120,6994	122,1271	,988309
8	148,0000	127,2500	116,3065	121,6356	121,6749	123,6990	,983636
9	136,0000	127,9583	106,2846	105,9819	128,3238	126,2854	1,016142
10	119,0000	128,5833	92,5470	92,1868	129,0857	128,3581	1,005669
11	104,0000	129,0000	80,6202	80,3857	129,3763	129,1879	1,001458
12	118,0000	129,7500	90,9441	90,1513	130,8910	130,6741	1,001660
13	115,0000	131,2500	87,6190	91,0641	126,2847	132,8723	,950421
14	126,0000	133,0833	94,6775	88,1204	142,9863	136,7653	1,045487
15	141,0000	134,9167	104,5090	100,8106	139,8662	137,3889	1,018031
16	135,0000	136,4167	98,9615	97,3073	138,7357	136,3299	1,017648
17	125,0000	137,4167	90,9642	98,2913	127,1731	133,8768	,949926
18	149,0000	138,7500	107,3874	111,4464	133,6966	134,5795	,993439
19	170,0000	140,9167	120,6387	122,6187	138,6412	137,6772	1,007001
20	170,0000	143,1667	118,7427	121,6356	139,7617	141,4112	,988336
21	158,0000	145,7083	108,4358	105,9819	149,0821	143,9746	1,035475
22	133,0000	148,4167	89,6126	92,1868	144,2723	145,5189	,991434
23	114,0000	151,5417	75,2268	80,3857	141,8163	148,0992	,957576
24	140,0000	154,7083	90,4929	90,1513	155,2944	153,6074	1,010983
25	145,0000	157,1250	92,2832	91,0641	159,2285	160,7892	,990293
26	150,0000	159,5417	94,0193	88,1204	170,2217	167,2294	1,017894
27	178,0000	161,8333	109,9897	100,8106	176,5687	171,0433	1,032305
28	163,0000	164,1250	99,3145	97,3073	167,5106	170,6210	,981770
29	172,0000	166,6667	103,2000	98,2913	174,9901	168,6987	1,037294
30	178,0000	169,0833	105,2735	111,4464	159,7180	164,9813	,968098
31	199,0000	171,2500	116,2044	122,6187	162,2917	164,6803	,985496
32	199,0000	173,5833	114,6423	121,6356	163,6034	166,4523	,982884
33	184,0000	175,4583	104,8682	105,9819	173,6146	171,4919	1,012378
34	162,0000	176,8333	91,6117	92,1868	175,7302	176,1563	,997581
35	146,0000	178,0417	82,0033	80,3857	181,6244	180,6664	1,005302
36	166,0000	180,1667	92,1369	90,1513	184,1348	185,6899	,991625

Prognozowanie

Pojęcie prognozy:

- „racjonalne, naukowe przewidywanie przyszłych zdarzeń”,
- „stwierdzenie odnoszącym się do określonej przyszłości formułowanym z wykorzystaniem metod naukowym, weryfikowalnym empirycznie, niepewnym, ale akceptowalnym”.

Ale

„*Prognozowanie to sztuka przewidywania przyszłości ... i uzasadniania, dlaczego owe przewidywania się nie sprawdzają.*”

O pewnych założeniach:

- zaobserwowany model nie zmieni się co do kształtu jak i siły działania w okresie przyszłym,
- wahania przypadkowe nie zakłócą znacząco zaobserwowanych składowych systematycznych modelu.

Spełnienie założeń bardziej prawdopodobne dla okresów leżących bliżej ostatniego badanego okresu.

Niezbędna wiedza dziedzinowa na temat charakteru zjawisk.

Pojęcia prognozowania

Okres prognozy

Horyzont prognozy

Zmiany ilościowe i jakościowe w prognozowaniu zjawisk.

Podział prognoz:

- krótkoterminowa,
- średnioterminowa,
- długoterminowa.

Ocena wiarygodności prognozy
szacowanie błędów prognozy *ex post*:

- bezwzględny błąd w momencie t: $y_t - y_t^*$,
- względny błąd w momencie t: $\frac{y_t - y_t^*}{y_t}$ 100%,
- średni (względny) błąd lub błąd kwadratowy.

Mining time streams (za J.Han, M.Kamber)

Similarity Search in Time-Series Analysis

- Normal database query finds exact match
- Similarity search finds data sequences that differ only slightly from the given query sequence

Data Transformations

Many techniques for signal analysis require the data to be in the frequency domain

- Usually data-independent transformations are used

The transformation matrix is determined a priori

  - discrete Fourier transform (DFT)
  - discrete wavelet transform (DWT)

- The distance between two signals in the time domain is the same as their Euclidean distance in the frequency domain