

Programowanie w języku R. Analiza szeregów czasowych

Aleksander Denisiuk
Uniwersytet Warmińsko-Mazurski
Olsztyn, ul. Słoneczna 54
denisjuk@matman.uwm.edu.pl

8 maja 2019

Analiza szeregów czasowych

Szeregi czasowe

W R

Prognoza

Najnowsza wersja tego dokumentu dostępna jest pod adresem

<http://wmii.uwm.edu.pl/~denisjuk/uwm>

Szeregi czasowe

- ❖ Szereg czasowy
- ❖ Analiza trendu
- ❖ Analiza sezonowości

W R

Prognoza

Szeregi czasowe

Szereg czasowy

Szeregi czasowe

❖ Szereg czasowy

❖ Analiza trendu

❖ Analiza
sezonowości

W R

Prognoza

- Ciąg obserwacji pokazujący kształtowanie się badanego zjawiska w kolejnych okresach
 - ◆ dniach, miesiącach, kwartałach, latach, itp.
- Składnik systematyczny
 - ◆ tendencja rozwojowa (trendem)
 - ◆ wahania cykliczne (składowa cykliczna)
 - wahania sezonowe (składowa sezonowa)
- zakłócenia losowe
- Główne cele:
 - ◆ wykrywanie natury zjawiska reprezentowanego przez sekwencję obserwacji
 - ◆ prognozowanie

Analiza trendu

Szeregi czasowe

❖ Szereg czasowy

❖ **Analiza trendu**

❖ Analiza
sezonowości

W R

Prognoza

- Wygładzanie
- Dopasowanie funkcji trendu

Analiza sezonowości

Szeregi czasowe

❖ Szereg czasowy

❖ Analiza trendu

❖ Analiza
sezonowości

W R

Prognoza

- Autokorelacja
 - ◆ korelacja między i -tym elementem szeregu a $(i+k)$ -tym elementem
 - k — opóźnienie
 - ◆ znaleźć k z maksymalną korelacją
- Korelacja wzajemna (cross-korelacja)

Szeregi czasowe

W R

❖ Typ danych

❖ Analiza

Prognoza

Szeregi czasowe w R

Typ danych

Szeregi czasowe

W R

❖ Typ danych

❖ Analiza

Prognoza

- Typ danych `ts`
- Przy wczytywaniu danych daty konwertować `as.Date()`

```
dates.df <- data.frame(dates=c("2019-04-18",  
                                "2019-04-19", "2019-04-20"))  
str(dates.df$dates)  
dates.1 <- as.Date(dates.df$dates, "%Y-%m-%d")  
str(dates.1)
```


Utworzenie szeregu

Szeregi czasowe

W R

❖ Typ danych

❖ Analiza

Prognoza

```
ts(1:10, frequency = 4, start = c(1959, 2))
```

- Co kwartał
- Poczynając od drugiego kwartału 1959
- Z macierzy:

```
z <- ts(matrix(rnorm(300), 100, 3),  
            start=c(1961, 1), frequency=12)  
class(z)
```

- Wizualizacja:

```
plot(z, plot.type="single", lty=1:3)
```

Przykład

Szeregi czasowe

W R

❖ Typ danych

❖ Analiza

Prognoza

- Dane `data/leaf2-4.txt` zawierają obserwacje liścia rosiczki. Co 40 minut, 36 obserwacji.
 - ◆ kolumna `FORM` — kształt: 1 — płaski, 2 — wklęsły
 - może być 1,5
- Wczytywanie danych i konwersja na szereg czasowy:

```
file.show("data/leaf2-4.txt")
leaf <- read.table("data/leaf2-4.txt",
  head=TRUE, sep=";")
str(leaf)
summary(leaf)
leaf.ts <- ts(leaf$FORM, frequency=36)
str(leaf.ts)
```

Autokorelacja i trend

Szeregi czasowe

W R

❖ Typ danych

❖ Analiza

Prognoza

● Autokorelacja

```
leaf.acf <- acf(leaf.ts)
leaf.acf
```

◆ autokorelacja i wykres (lag — opóźnienie)

■ niebieska linia: poziom ufności 95%

◆ w jakich przedziałach czasowych korelacja jest istotna?

● Trend, składowa sezonowa i losowa

```
leaf.stl <- stl(leaf.ts, s.window="periodic")
leaf.stl
plot(leaf.stl$time.series)
```

◆ s.window — nieparzysta liczba powyżej 6, albo "periodic" — okres wahań sezonowych

Szeregi czasowe

W R

Prognoza

- ❖ Dane
- ❖ Model
- ❖ Dobór order
- ❖ Prognoza
- ❖ forecast

Prognoza

Szeregi czasowe

W R

Prognoza

❖ Dane

❖ Model

❖ Dobór order

❖ Prognoza

❖ forecast

- Dane o ilości nowych abonentów operatora internetowego od grudnia 2004 do 2008 roku

```
new.users <-  
  ts(read.table("data/data.txt")$V3,  
      start=c(2004,12), frequency=12)  
tot.users <- ts(cumsum(new.users),  
               start=c(2004,12), frequency=12)  
oldpar <- par(mfrow=c(2,1))  
plot(new.users, type="b", log="y", xlab="")  
plot(tot.users, type="b", ylim=c(1,3000),  
      log="y")  
  
par(oldpar)
```

Model

Szeregi czasowe

W R

Prognoza

❖ Dane

❖ Model

❖ Dobór `order`

❖ Prognoza

❖ `forecast`

- Metoda ARIMA (Autoregressive Integrated Moving Average)
- `arima(ts, order)`
 - ✦ `parametr order = c(p, d, q)`
 - `p` — opóźnienie autoregresji (lag)
 - `d` — rząd ilorazu różniczkowego
 - `q` — stopień regresji
- Dobór najlepszego parametru `order`
 - ✦ kryterium AIC
 - ✦ pakiet `forecast`

Dobór q

Szeregi czasowe

W R

Prognoza

❖ Dane

❖ Model

❖ Dobór order

❖ Prognoza

❖ forecast

```
model01 <- arima(tot.users, order=c(0,0,1))
model02 <- arima(tot.users, order=c(0,0,2))
model03 <- arima(tot.users, order=c(0,0,3))
```

.....

```
model012 <- arima(tot.users, order=c(0,0,12))
model013 <- arima(tot.users, order=c(0,0,13))
model014 <- arima(tot.users, order=c(0,0,14))
```

```
plot(AIC(model01, model02, model03, model04,
          model05, model06, model07, model08, model09,
          model010, model011, model012, model013,
          model014), type="b")
```

- jaka jest wartość optymalna?

Dobór q w pętli

Szeregi czasowe

W R

Prognoza

❖ Dane

❖ Model

❖ Dobór order

❖ Prognoza

❖ forecast

```
for (m in 1:14) {  
  assign( paste("model0",m, sep=""),  
    arima(tot.users, order=c(0,0,m)) )  
}
```

- `assign` zastępuje operator `<-`
- optymalna wartość $q = 12$

Dobór p

Szeregi czasowe

W R

Prognoza

❖ Dane

❖ Model

❖ Dobór order

❖ Prognoza

❖ forecast

```
model012 <- arima(tot.users, order=c(0,0,12))  
model112 <- arima(tot.users, order=c(1,0,12))  
model212 <- arima(tot.users, order=c(2,0,12))  
model312 <- arima(tot.users, order=c(3,0,12))  
model412 <- arima(tot.users, order=c(4,0,12))
```

```
AIC(model012, model112, model212,  
      model312, model412)
```

- jaka jest wartość optymalna?

Dobór d

Szeregi czasowe

W R

Prognoza

❖ Dane

❖ Model

❖ Dobór order

❖ Prognoza

❖ forecast

```
model1120 <- arima(tot.users, order=c(1,0,12))
model1121 <- arima(tot.users, order=c(1,1,12))
model1122 <- arima(tot.users, order=c(1,2,12))
model1123 <- arima(tot.users, order=c(1,3,12))
model1124 <- arima(tot.users, order=c(1,4,12))
model1125 <- arima(tot.users, order=c(1,5,12))
AIC(model1120, model1121, model1122, model1123,
     model1124, model1125)
```

- jaka jest wartość optymalna?

Prognoza

Szeregi czasowe

W R

Prognoza

❖ Dane

❖ Model

❖ Dobór order

❖ Prognoza

❖ forecast

- Funkcja `predict()`

```
pred.users <- predict(model1123, n.ahead=12,  
                        se.fit = TRUE)
```

- ◆ prognoza na 12 miesięcy
- ◆ oblicza się błąd standardowy (`se.fit = TRUE`)

- Wykres

```
plot(tot.users, xlim=c(2004.7, 2010),  
     ylim=c(0, 6500))  
lines(pred.users$pred, col="green")  
lines(pred.users$pred + pred.users$se,  
      col="red")  
lines(pred.users$pred - pred.users$se,  
      col="red")
```

Pakiet forecast

Szeregi czasowe

W R

Prognoza

❖ Dane

❖ Model

❖ Dobór order

❖ Prognoza

❖ **forecast**

```
library(forecast)
model <- auto.arima(tot.users)
```