

# Wprowadzenie

Analiza szeregów czasowych zyskuje ostatnio coraz bardziej na znaczeniu i jest z niesłabnącym powodzeniem stosowana w wielu obszarach nauki, biznesu czy przemysłu. Podstawowym celem analizy szeregów czasowych jest zbudowanie modelu, który będzie dobrze opisywał dynamikę czasową obserwowanego zjawiska i który może być następnie wykorzystany do prognozowania przyszłych (nieznanych) wartości. Przed właściwym modelowaniem konieczne jest oczywiście odpowiednie przygotowanie danych, a w szczególności uwzględnienie na tym etapie sposobu gromadzenia danych i wykonanie przekształceń, które mogą ułatwić dopasowanie modelu. Duże znaczenie ma również prawidłowa identyfikacja regularności występujących w analizowanym szeregu, takich jak tendencje długoterminowe (trendy) oraz wahania sezonowe (sezonowość).

Podobnie jak w przypadku innych nowoczesnych metod analizy ilościowej, duże znaczenie dla rozwoju metod analizy i prognozowania szeregów czasowych miał rozwój technologii informatycznych. Istotny dla upowszechnienia metod analizy szeregów jest łatwy dostęp do specjalistycznego oprogramowania, dającego praktykom możliwość zastosowania zarówno standardowych, jak i nowoczesnych algorytmów oraz przeprowadzenia zaawansowanych obliczeń. Szczególną rolę odegrał w tym przypadku rozwój darmowego oprogramowania, takiego jak system statystyczny **R** ([www.r-project.org](http://www.r-project.org)). Dzięki temu dostęp do metod analizy i prognozowania szeregów stał się powszechny, a nowoczesne metody opracowywane w ośrodkach naukowych są często w krótkim czasie udostępniane w postaci dodatkowych pakietów (bibliotek), rozszerzających standardowe możliwości środowiska **R**.

Pierwszy rozdział książki ma charakter wprowadzający i omawiamy w nim najważniejsze zagadnienia związane z analizą szeregów czasowych. Będzie mowa m.in. o tym, czym jest szereg czasowy i jakie są podstawowe zadania analizy szeregów. Szczególną uwagę poświęcamy przeglądowi etapów występujących w analizie szeregów czasowych, począwszy od niezbędnych przekształceń, poprzez identyfikację regularnych wzorców, a kończąc na prognozowaniu.

Co ważne, ten rozdział zawiera „przykład dla niecierpliwych”, który pokazuje, jak wykonać kompletną analizę szeregu czasowego w **R**. Przykład jest zaprezentowany w możliwie uproszczony sposób, aby ułatwić zrozumienie procesu. Zawiera też odnośniki do rozdziałów, w których poszczególne zagadnienia są szerzej omówione.

Zapoznanie się z tym wprowadzającym rozdziałem ułatwi Czytelnikowi orientację w strukturze książki i pozwoli zrozumieć całość procesu analizy szeregu czasowego.

## 2.1. Czym jest szereg czasowy?

Szeregiem czasowym (ang. *time series*) nazywamy obserwacje interesującej nas wielkości zarejestrowane w kolejnych (zazwyczaj regularnych) odstępach czasu, np. kolejnych dniach, miesiącach lub kwartałach. Przykładami mogą być szeregi zawierające informacje o rocznej wielkości produkcji samochodów osobowych, stopie bezrobocia w kolejnych miesiącach czy też kwartalne dane dotyczące przyrostu produktu krajowego brutto (PKB).

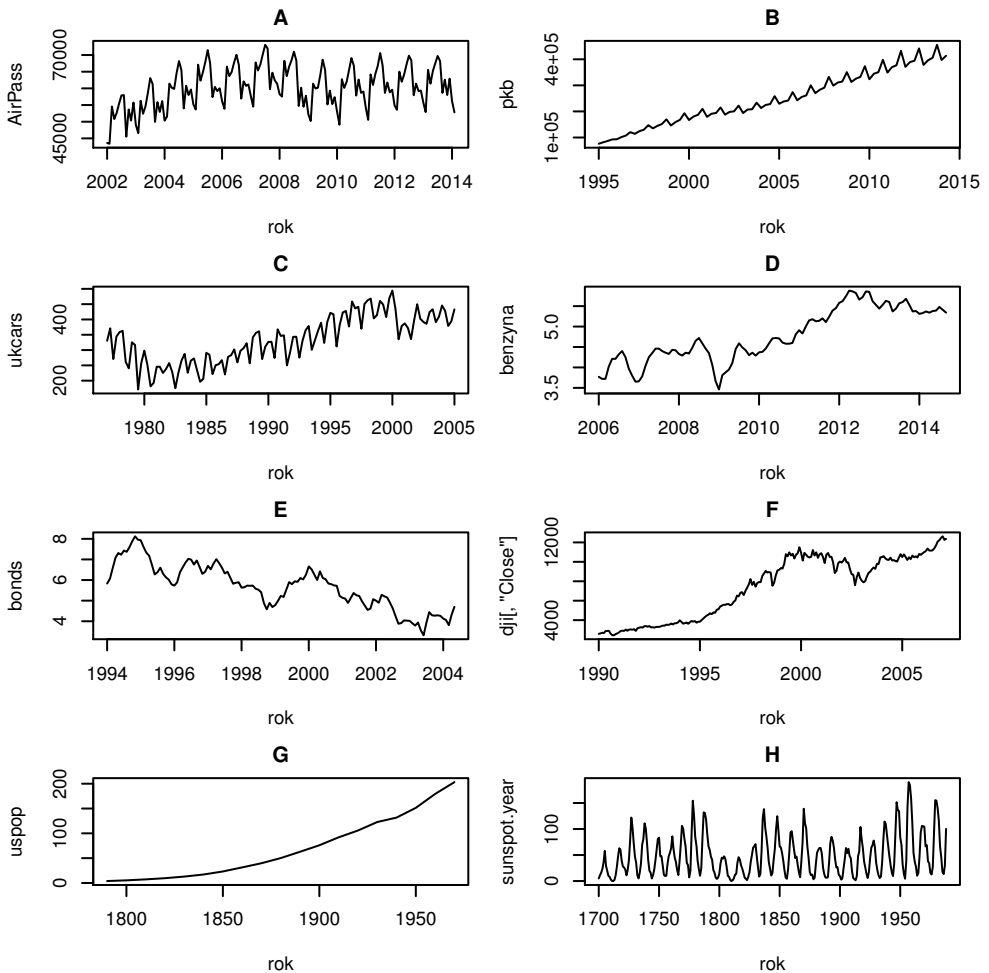
Rysunek 2.1 przedstawia typowe szeregi czasowe z różnych obszarów zastosowań. Mamy następujące dane<sup>1</sup>:

- szereg A:** miesięczna liczba pasażerów linii lotniczych (w tysiącach) w USA, w latach 2002–2014,
- szereg B:** kwartalne wartości produktu krajowego brutto (PKB) w Polsce, zarejestrowane w okresie 1995–2014,
- szereg C:** kwartalna wielkość produkcji samochodów osobowych w UK w okresie 1977:1–2005:1,
- szereg D:** miesięczne, średnie ceny 1 litra benzyny w Polsce w okresie 2006–2014,
- szereg E:** rentowność 10-letnich obligacji skarbowych USA, dane miesięczne w okresie styczeń 1994–maj 2004,
- szereg F:** miesięczne kursy zamknięcia indeksu Dow Jones w okresie styczeń 1990–marzec 2007,
- szereg G:** populacja USA (w milionach) w okresie 1790–1970, dane 10-letnie,
- szereg H:** roczne liczby plam słonecznych w latach 1700–1988.

Jak widzimy, szeregi czasowe mogą różnić się częstotliwością próbkowania, czyli interwałami czasowymi pomiędzy kolejnymi obserwacjami (np. kolejne dni, miesiące, kwartały, lata<sup>2</sup> itd.).

<sup>1</sup>Dokładniejszą informację na temat źródła pochodzenia prezentowanych danych oraz ich dostępności w środowisku **R** przedstawimy w podrozdziale 3.1.

<sup>2</sup>W podręczniku ograniczamy się do tzw. regularnych szeregów czasowych, które charakteryzują się jednakowymi odstępami czasowymi pomiędzy kolejnymi obserwacjami.



**Rysunek 2.1.** Przykłady szeregów czasowych z różnych obszarów zastosowań

Analizując wykresy przykładowych szeregów (rys. 2.1), możemy zauważyć duże zróżnicowanie pod względem występujących regularności. W szczególności, mamy szeregi, w których obecna jest wyraźna tendencja długoterminowa (trend) oraz takie, w przypadku których łatwo można dostrzec zachowania okresowe (sezonowe).

Kolejne obserwacje szeregu czasowego charakteryzują się zatem nieprzypadkowym porządkiem i wykazują najczęściej istotną zależność (korelację). To właśnie badanie charakteru i siły tej zależności jest podstawowym zadaniem analizy szeregów czasowych. Oczywiście zależność pomiędzy obserwacjami jest często wykorzystywana do prognozowania przyszłych wartości szeregu.

Dla szeregów czasowych stosowane są zwykle oznaczenia:  $X_t$ ,  $Y_t$  lub  $Z_t$ , w których argument  $t$  oznacza czas (ang. *time*) i należy do ustalonego zbioru

chwil (punktów czasowych), np.  $t \in I$ . Często zamiast operowania rzeczywistą skalą czasową (daty kalendarzowe), obserwacje indeksuje się kolejnymi liczbami naturalnymi. Dla przykładu szereg zawierający dane miesięczne dla kolejnych 12 lat (łącznie 144 obserwacje) zapisujemy dla uproszczenia jako  $X_1, X_2, \dots, X_{144}$ .

## 2.2. Główne zadania analizy szeregów czasowych

Z szeregami czasowymi często się spotykamy, gdy zachodzi konieczność podejmowania ważnych decyzji biznesowych, np. dotyczących kupna/sprzedaży, produkcji, zatrudnienia czy logistyki. Na początek wymienimy kilka typowych przykładów analiz, w których wykorzystuje się szeregi czasowe:

- ⇒ Prognozowanie wielkości sprzedaży lub popytu na określony produkt/surowiec w kolejnych okresach.
- ⇒ Prognozy wartości wskaźników makroekonomicznych (np. inflacji lub PKB) w kolejnych kwartałach.
- ⇒ Analiza sytuacji na rynku pracy (w szczególności analiza tendencji dotyczących bezrobocia i zatrudnienia, w różnych grupach wiekowych).
- ⇒ Prognozowanie wartości akcji danej spółki, cen surowców, kursów walutowych itp. w kolejnych okresach.
- ⇒ Przewidywanie zmian cen danego produktu (np. paliw) w kolejnych miesiącach.
- ⇒ Analiza zmian demograficznych, socjologicznych, klimatycznych i ich wpływu na koniunkturę w określonej gałęzi przemysłu.

Wymienione przykładowe zastosowania analizy szeregów czasowych cechuje więc dość duża różnorodność. Wylaniają się tutaj jednak dwa główne zadania, tzn. identyfikacja regularnych tendencji (tzw. dekompozycja szeregu czasowego) oraz prognozowanie. Regularne tendencje to przede wszystkim trend – długoterminowa tendencja rozwojowa, np. wzrostowa lub spadkowa, oraz sezonowość – cykliczne wahania wartości szeregu wokół tendencji rozwojowej, związane najczęściej z danym miesiącem roku, porami wakacji lub zmianami pogody.

Przed analitykiem, wyciągającym wnioski na podstawie analizy szeregów czasowych, pojawia się więc wiele pytań, w tym na przykład:

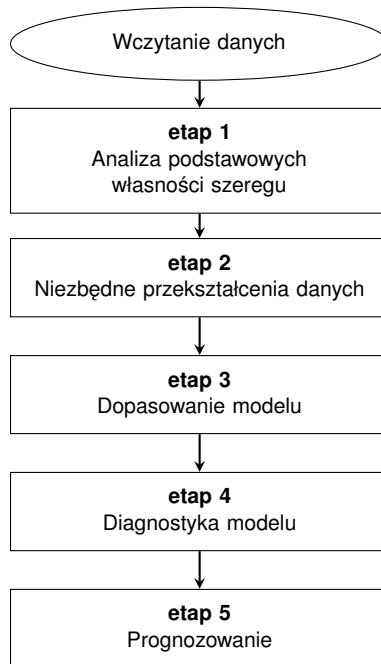
- ⇒ Jak przygotować dane przed właściwą analizą?
- ⇒ Jakie metody analizy i prognozowania szeregów powinny być zastosowane?
- ⇒ Jak mierzyć błąd prognozy i czy uzyskane prognozy są wystarczająco dokładne?

W kolejnych rozdziałach naszego podręcznika postaramy się odpowiedzieć na te i podobne pytania.

## 2.3. Etapy w analizie szeregu czasowego

Analiza szeregu czasowego jest zazwyczaj procesem wieloetapowym. Rysunek 2.2 przedstawia typowe etapy w analizie szeregu czasowego. Niektóre etapy są opcjonalne i mogą być ściśle związane ze specyfiką danych, na przykład sposobem rejestrowania danych, regularnościami występującymi w analizowanym szeregu itp. Większość metod i modeli stosowanych w analizie szeregów występuje zazwyczaj w wielu wariantach lub wymaga wybrania adekwatnych parametrów. Po przeprowadzeniu określonego etapu analizy szeregu czasowego może się okazać, że konieczny będzie powrót do poprzednich etapów. Dla przykładu, jeżeli przeprowadzając diagnostykę (etap 4) stwierdzimy, że model nie jest dobrze dopasowany do danych, możemy powrócić do etapu dopasowania (etap 3) i spróbować znaleźć lepszy model dla naszego szeregu czasowego. Niekiedy konieczny może być także powrót do etapu 2 i wykonanie dodatkowych przekształceń, które mogą poprawić jakość dopasowania modelu. Oczywiście, może się również zdarzyć tak, że konieczne będzie powtórne przyjrzenie się podstawowym własnościom analizowanego szeregu, a więc powrót do etapu 1.

Wykonanie kolejnych etapów analizy szeregu czasowego wymaga zarówno dobrej znajomości samej metodologii, jak i odpowiedniego doświadczenia



Rysunek 2.2. Ogólny schemat analizy szeregów czasowych

analityka. Z pomocą przychodzą tu rozwiązania (m.in. odpowiednie funkcje dostępne w środowisku **R**) opracowane z myślą o automatyzacji pewnych etapów analizy. Stosowanie tego rodzaju narzędzi nie zwalnia oczywiście analityka z konieczności weryfikacji uzyskanych wyników.

Poszczególne etapy w analizie szeregu czasowego (w szczególności stosowane modele i metody) będziemy sukcesywnie omawiali w kolejnych rozdziałach książki. Na wstępie chcielibyśmy jednak umożliwić Czytelnikowi spojrzenie na całość tego procesu, przy okazji prezentując pokrótce (bez nadmiernych szczegółów technicznych) możliwości oferowane w tym zakresie w pakiecie **R**. Z tego względu w podrozdziale 2.4 przedstawiona jest możliwie kompletna analiza wybranego szeregu czasowego, z uwzględnieniem najbardziej popularnych metod i modeli.

## 2.4. Przykład dla niecierpliwych

Na potrzeby przykładu wybraliśmy szereg czasowy zawierający informacje o liczbie turystów korzystających z noclegów na terenie Dolnego Śląska. Są to dane miesięczne, zarejestrowane w okresie styczeń 2009–marzec 2014. Dane zostały pobrane ze strony GUS: Bank Danych Lokalnych, dział TURYSTYKA (<http://stat.gov.pl/bd1/>).

Analiza, którą przeprowadzimy, będzie obejmowała następujące kroki:

1. Wczytanie (import) danych do środowiska **R**.
2. Konwersja danych na format odpowiedni dla **R**.
3. Analiza podstawowych własności szeregu.
4. Dekompozycja: identyfikacja trendu i sezonowości.
5. Niezbędne przekształcenia szeregu.
6. Podział danych na zbiór uczący i testowy.
7. Dopasowanie modeli.
8. Prognozowanie: konstrukcja prognoz punktowych i przedziałów predykcyjnych.
9. Ocena i porównanie dokładności prognoz.

### 2.4.1. Wczytanie danych

Zakładamy, że kolejne wartości szeregu zapisane są w pliku tekstowym `hotele.txt`<sup>3</sup>, który został umieszczony w wybranym katalogu (w naszym przypadku, w katalogu `C:/Users/adam/Desktop/Dane`). Poniższy fragment kodu pozwala wczytać dane do przestrzeni roboczej **R**'a.

---

<sup>3</sup>Plik `hotele.txt` dostępny jest na stronie <http://TSAFBook.quantup.pl>. Dane w formacie binarnym **R**'a można także znaleźć w towarzyszącym książce pakiecie **TSAFBook**.