Analiza danych rzeczywistych przy pomocy modelu ARMA

Natalia Klepacka, Joanna Kołaczek

8 lutego 2023

Spis treści

1 Wstęp **2**

2 Przygotowanie danych do analizy **2**

3 Modelowanie danych przy pomocy ARMA **5**
3.1 Dobranie rzędu modelu 5
3.2 Estymacja parametrów modelu 6

4 Ocena dopasowania modelu **6**

5 Weryfikacja założeń dotyczących szumu **6**
5.1 Średnia 6
5.2 Wariancja 7
5.3 Niezależność 7
5.4 Normalność rozkładu 7

6 Podsumowanie **7**

7 Źródła **7**

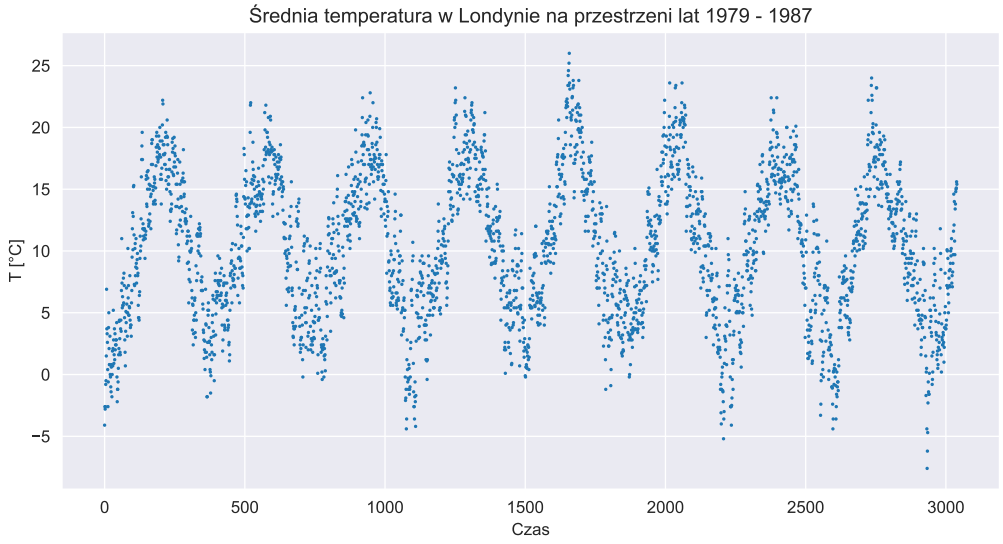
1 Wstęp

Niniejszy raport powstał na potrzeby realizacji laboratorium z Komputerowej Analizy Szeregów Czasowych, prowadzonych przez mgr Justynę Witulską, do wykładu prof. Agnieszki Wyłomańskiej. Będziemy analizować dane dotyczące średniej dziennej temperatury w Londynnie, na przestrzeni lat 1979-2021. Dysponujemy 15304 obeserwacjami, jednak chcąc aby analiza była bardziej precyzyjna, będziemy rozważać tylko pierwszych 3040. Dane pochodzą *z tej strony*. Są to informacje kolekcjonowane przez European Climate Assessment and Dataset - projekt zbierający dane o pogodzie w Europie. W raporcie przyjrzymy się przebiegowi dekompozycji Walda, oraz przy pomocy kryteriów informacyjnych dobierzemy rzędy modelu ARMA aby następnie wyestymować wartości parametrów tegoż modelu. Zweryfikujemy również, czy założenia dotyczące szumu się zgadzają.

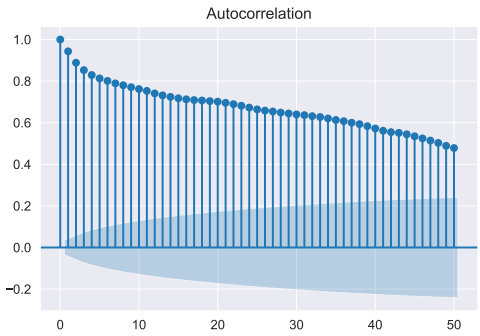
Życzymy Czytelnikowi miłej lektury.

2 Przygotowanie danych do analizy

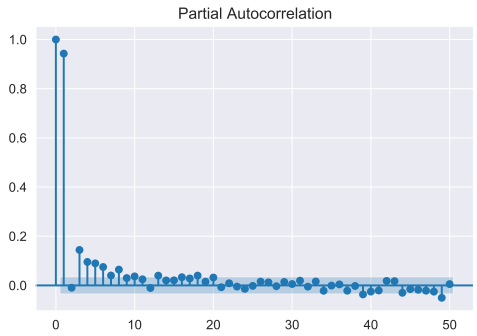
Na wykresie [1] przedstawiona jest zależność średniej dziennej temperatury w Londynie, w zależności od czasu. Wyraźnie widoczna jest sezonowość, natomiast nie możemy być pewni co do obecności trendu. Wykresy autokorelacji [2] oraz częściowej autokorelacji [3] potwierdzają, że nie możemy mówić tu o szeregu stacjonarnym.



Rysunek 1: Wykres temperatury w Londynie

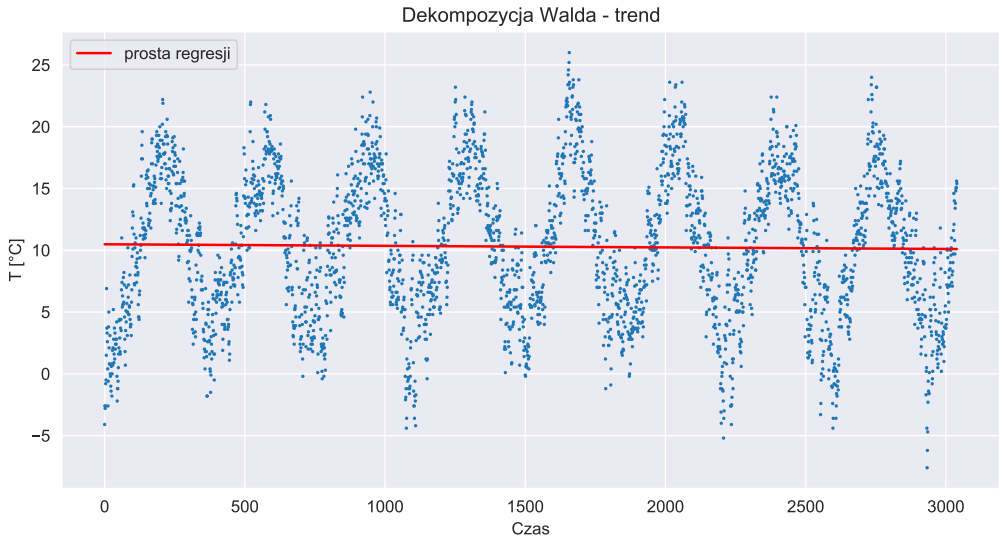


Rysunek 2: Autokorelacja



Rysunek 3: Częściowa autokorelacja

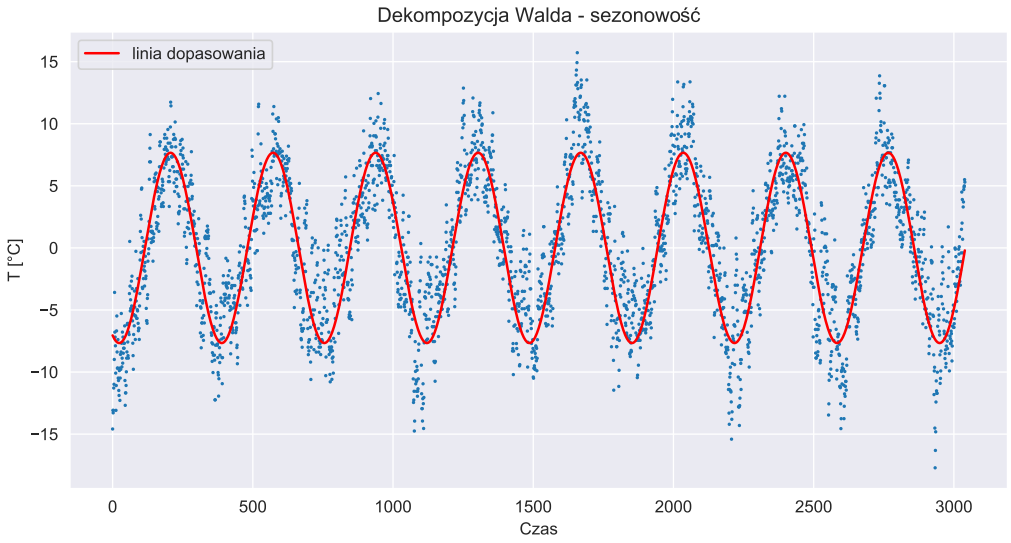
Aby usunąć możliwy trend, stosujemy dla naszych danych regresję liniową. Efekt widzimy na wykresie [4]. Obliczony współczynnik kierunkowy był bardzo bliski zeru co oznacza, że w danych nie występuje trend liniowy, natomiast wyraz wolny wyniósł około 10, odejmujemy tą wartość od wartości oryginalnych aby je ustandaryzować. (?????????/)



Rysunek 4: Regresja liniowa

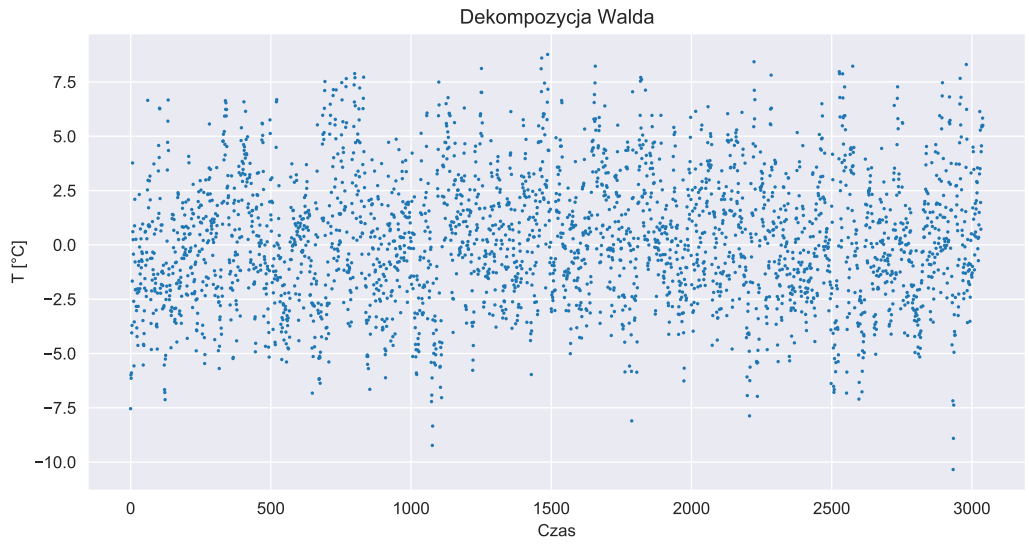
Kolejnym krokiem będzie próba usunięcia sezonowości. Na poprzednich wykresach mogliśmy zauważyć, że dane układają się w kształt funkcji sinusoidalnej. Załóżmy, że da się je opisać przy pomocy funkcji $f(x) = c \cdot \sin(d \cdot x + e)$. Używamy pakietu `scipy`, a konkretnie funkcji `optimize.curve_fit`,

aby dopasować odpowiednie współczynniki c, d, e. Efekt widzimy na wykresie 5.

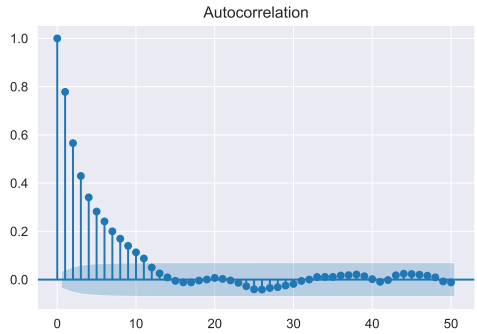


Rysunek 5: Krzywa sinusoidalna

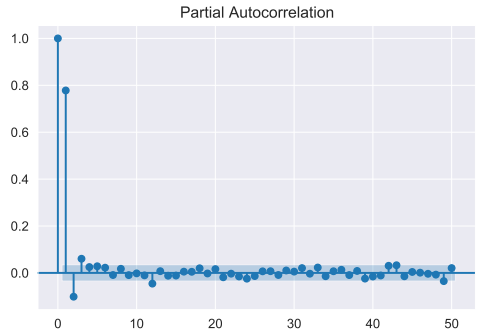
Aby dokończyć proces dekompozycji, wystarczy od naszych danych odjąć wartości otrzymanej funkcji w danym czasie [6]. Możemy teraz ponownie sprawdzić jak prezentują się wykresy autokorelacji i częściowej autokorelacji. .....



Rysunek 6: Dane po dekompozycji



Rysunek 7: Autokorelacja



Rysunek 8: Częściowa autokorelacja

3 Modelowanie danych przy pomocy ARMA

3.1 Dobranie rzędu modelu

W celu dobrania rzędu modelu, wyliczyłyśmy wartości kilku kryteriów informacyjnych

- Kryterium Informacyjne Akaikego (AIC),
- Bayesowskie Kryterium Informacyjne (BIC),
- Kryterium Informacyjne Hannana-Quinna (HQIC)

dla wartości p i q z przedziału $[0, 9]$. Otrzymałyśmy wyniki jak w tabelach poniżej. Najlepiej dopasowane pary wg poszczególnych kryteriów to

- $p = 5$, $q = 6$ dla AIC,
- $p = 1$, $q = 1$ dla BIC,
- $p = 3$, $q = 0$ dla HQIC.

Jako że nie da się określić modelu jednoznacznie najlepiej dopasowanego, do dalszej analizy zdecydowałyśmy się przyjąć najmniej skomplikowany z powyższych modeli, czyli ARMA(1, 1).

3.2 Estymacja parametrów modelu

Do estymacji parametrów wykorzystamy metodę największej wiarygodności zaimplementowaną w pythonowym pakiecie statsmodels. Metoda ta zakłada, że zmienne składające się na szum mają rozkład normalny. Wartości współczynników można zobaczyć w tabeli poniżej.

4 Ocena dopasowania modelu

5 Weryfikacja założeń dotyczących szumu

Szum Z_t w modelu ARMA powinien spełniać poniższe założenia:

1. Średnia Z_t jest bliska 0
2. Wariancja Z_t jest taka sama dla każdego t
3. Z_t są niezależne
4. Z_t mają rozkład normalny

5.1 Średnia

Na podstawie wykresu możemy stwierdzić, że średnia prawdopodobnie jest równa 0. Dodatkowo możemy to sprawdzić przy pomocy testu t. Test ten sprawdza hipotezę zerową "Średnia z próby jest równa μ" przeciwko hipotezie alternatywnej "Średnia z próby jest różna od μ". Jak widać poniżej, wynik testu potwierdza wnioski wyciągnięte z wykresu — nie ma podstaw do odrzucenia zerowej.

5.2 Wariancja

Na wykresie nie widać znaczących zmian w wariancji zależnych od czasu, jednak dla pewności wykonamy test

5.3 Niezależność

Z wykresów funkcji autokorelacji i częściowej autokorelacji można łatwo wywnioskować, że residua są realizacjami niezależnych zmiennych losowych.

5.4 Normalność rozkładu

Dystrybuanta rozkładu residuów niemal idealnie pokrywa się z dystrybuantą rozkładu normalnego. Podobnie histogram wartości dobrze przybliża teoretyczną gęstość. Wykres kwantylowy pokrywa się odrobinę gorzej, ale nadal wystarczająco dobrze, żebyśmy mogli uznać, że dane pochodzą z rozkładu normalnego.

6 Podsumowanie

Na podstawie przeprowadzonej analizy możemy stwierdzić, że między poziomem szczęścia a PKB per capita dla danego kraju występuje dosyć silna zależność liniowa. Niestety badanie residuów wykazało, że nie mają one rozkładu normalnego. Oznacza to, że co prawda punktowe estymacje zostały przez nas wykonane poprawnie, jednak wszystkie estymacje przedziałowe oparte były o założenie o normalności rozkładu residuów, zatem nie mają zastosowania w tym modelu. Niestety nie znając rozkładu ϵ, nie możemy wyznaczyć faktycznych przedziałów ufności.

7 Źródła

- Wykłady
- `https://www.kaggle.com/datasets/eliasturk/world-happiness-based-on-cpi-20152020`
- `https://www.itl.nist.gov/div898/handbook/eda/section3/eda35a.html`
- `http://tofesi.mimuw.edu.pl/~cogito/smarterpoland/samouczki/testyNormalnosci/testyNormalnosci.pdf` str.11