

Analiza danych rzeczywistych przy pomocy modelu ARMA

Natalia Klepacka, Joanna Kołaczek

08.02.2023

Spis treści

1	Wstęp	2
2	Przygotowanie danych do analizy	2
3	Ocena dopasowania modelu	5
4	Weryfikacja założeń dotyczących szumu	6
5	Podsumowanie	6
6	Źródła	6

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 poziomu szczęścia w wybranych krajach na świecie oraz jego związku z wartością PKB na osobę (w dalszej części raportu, będziemy je określać skrótowo jako szczęście i PKB). Po usunięciu wartości brakujących dysponujemy próbami o wielkości 791. Dane pochodzą z *tej strony*. Są to wyniki uzyskane przez Instytut Gallupa, w ankietach badających poziom szczęścia oraz jego możliwe indykatory, zebrane w latach 2015-2020. W raporcie przeprowadzimy analizę jednowymiarową dla dwóch zmiennych oraz zwizualizujemy je przy pomocy histogramu, dystrybuantry empirycznej oraz boxplotu. Następnie wyestymujemy współczynniki w klasycznym modelu regresji, aby ostatecznie sprawdzić, czy uzyskane residua spełniają oczekiwane założenia.

Życzymy Czytelnikowi miłej lektury.

2 Przygotowanie danych do analizy

3 Modelowanie danych przy pomocy ARMA

3.1 Dobranie rzędu modelu

W celu dobrania rzędu modelu, wyliczyliśmy wartości kilku kryteriów informacyjnych

- Kryterium Informacyjne Akaikiego (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

5.4 Normalność rozkładu

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