

# Projekt z Szeregów Czasowych

Jan Moskal

2025-01-18

## Wstęp

Dane, które będziemy analizować, pochodzą ze strony Głównego Urzędu Statystycznego (<https://bdl.stat.gov.pl/bdl/dane/podgrup/temat>) znajdują się w grupie “Przeciętne ceny detaliczne towarów i usług konsumpcyjnych”, podgrupie “Ceny detaliczne wybranych towarów i usług konsumpcyjnych (dane miesięczne)” i dotyczą cen węgla kamiennego za toną. Dane o przeciętnych cenach obejmują notowania co miesiąc dla całej Polski. Projekt ma na celu analizę tego szeregu czasowego, aby zrozumieć zmiany cen węgla kamiennego w Polsce w latach 2006-2019 i stworzyć prognozy na przyszłość.

## Wczytywanie danych

```
dane <- read_excel("wieg_kamienny_szereg.xlsx", range = "TABLICA!C4:FN6")  
  
dane <- as.vector(dane[2, ])  
dane <- as.numeric(unlist(dane))
```

Zamieniamy wektor w macierz, aby ustawić dobrą kolejność danych (nawet mamy dane wypisane, w ten sposób, że jeden miesiąc dla czternastu lat i dopiero następny miesiąc, a chcemy żeby było chronologicznie)

```
macierz <- matrix(dane, ncol = 14, byrow = TRUE)
```

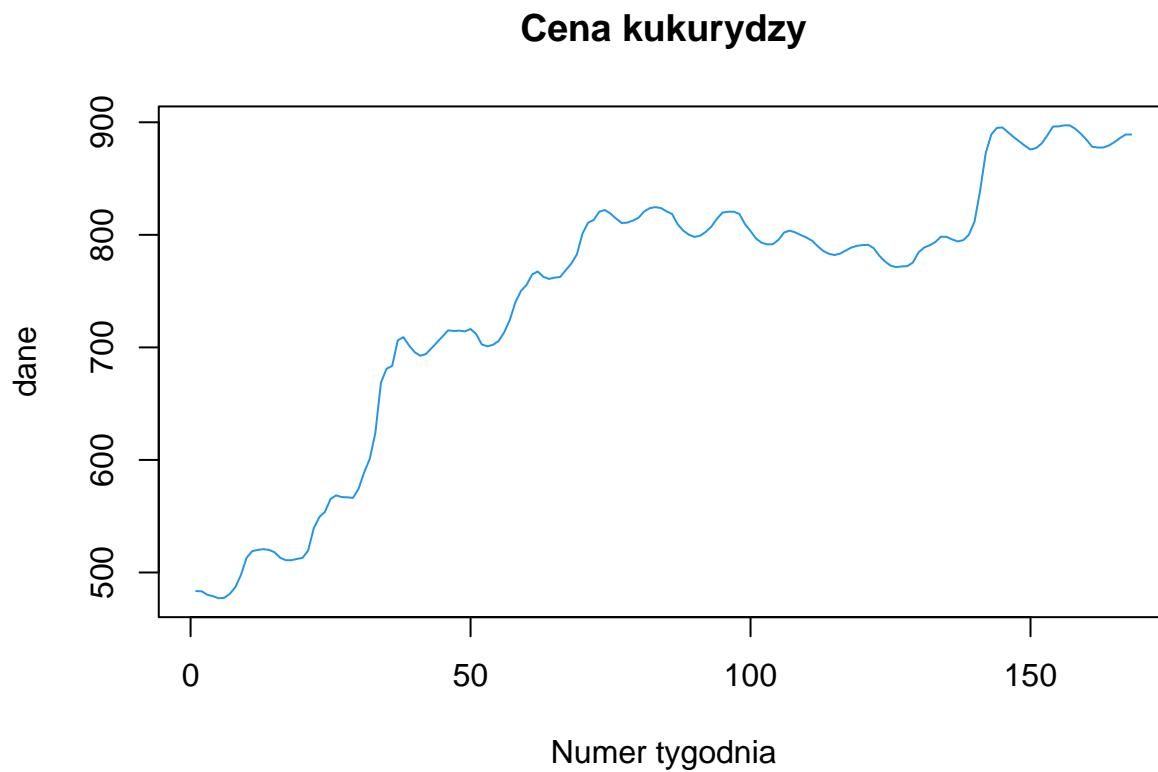
Przekształcenie macierzy w wektor czytany kolumnowo (od góry do dołu). W ten sposób otrzymujemy dane w odpowiedniej kolejności.

```
dane <- as.vector(macierz)  
t <- 1:length(dane)
```

## Wstępna analiza szeregu

Wykres liniowy dla naszych danych w czasie t.

```
plot(y=dane,x=t,col=4, main = "Cena kukurydzy", type = "l", xlab = "Numer tygodnia" )
```

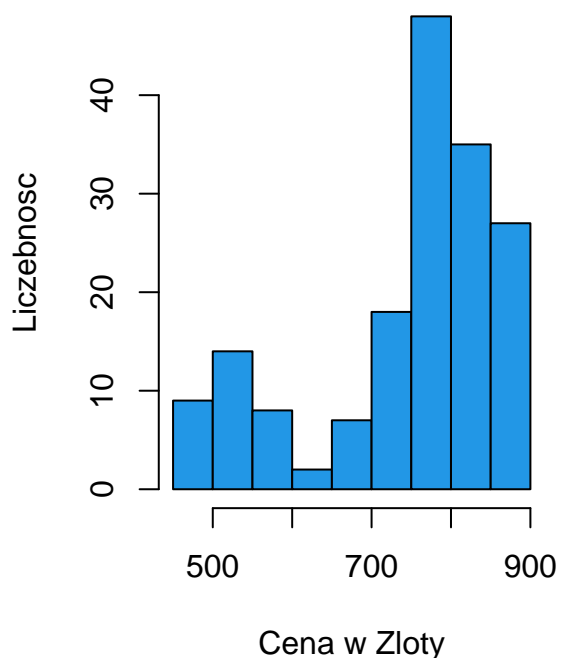


Jak widzimy z wykresu, cena dość szybko wzrosła do cen powyżej 700 zł. Widzimy również, że ogólny trend jest rosnący.

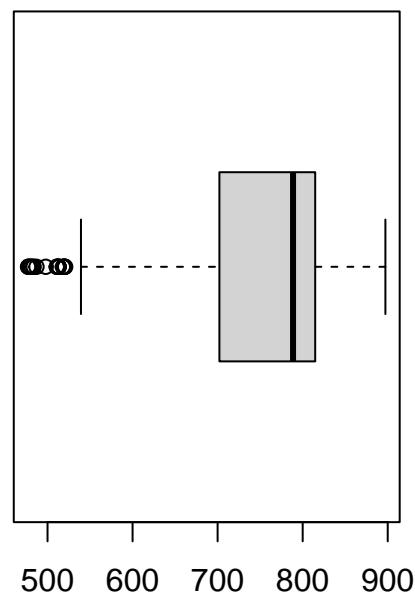
Robimy wykresy typu boxplot oraz histogram, żeby zobaczyć rozkład danych.

```
par(mfrow = c(1, 2))
hist(dane, breaks = 15, col = 4, main = "Histogram cen węgla kamiennego",
     xlab = "Cena w Złoty", ylab = "Liczebność", prob = FALSE)
boxplot(dane, col = "lightgrey", main = "Wykres ramka-wąsy", horizontal = TRUE)
```

## Histogram cen węgla kamiennego



## Wykres ramka-wasy



```
par(mfrow = c(1, 1))
```

Możemy zauważyć, że rozkład jest lewostronnie asymetryczny, bierzemy się to z tego co już zauważyliśmy z wykresu liniowego czyli, że ceny od 700 zł za tonę zaczęły się już po 2 latach od pierwszej obserwacji z szeregu a pozostałe 12 lat oscylowało co do wartości od 700 do 900 zł za tonę. Z wykresu pudełkowego możemy zauważyć nawet dokładniej, że kwartył pierwszy wynosi około 700 a kwartył trzeci około 810 co w przełożeniu na nasz problem oznacza, że połowa obserwacji, czyli z 7 lat znajduje się na tym małym przedziale.

Podstawowe statystyki

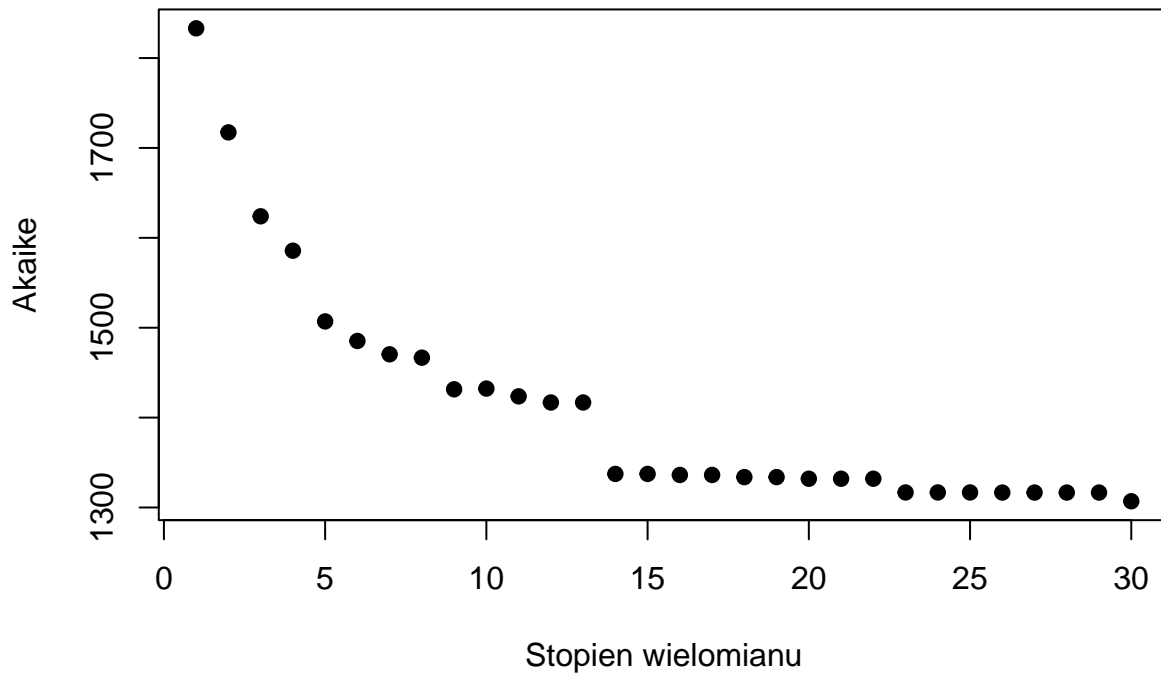
```
summary(dane)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  477.1   702.2   788.7   744.1   814.5   897.3
```

Szukamy najlepszego wielomianu opisującego nasz szereg

```
Akaike <- c()
for(i in 1:30){
  Akaike <- cbind(Akaike, dopasowanie_wielomianu(dane,i))
}
i=1:30
plot(i, Akaike,type="p", pch=19, main = "Kryterium AIC dla wielomianu stopnia i", xlab = "Stopień wielomianu")
```

## Kryterium AIC dla wielomianu stopnia i

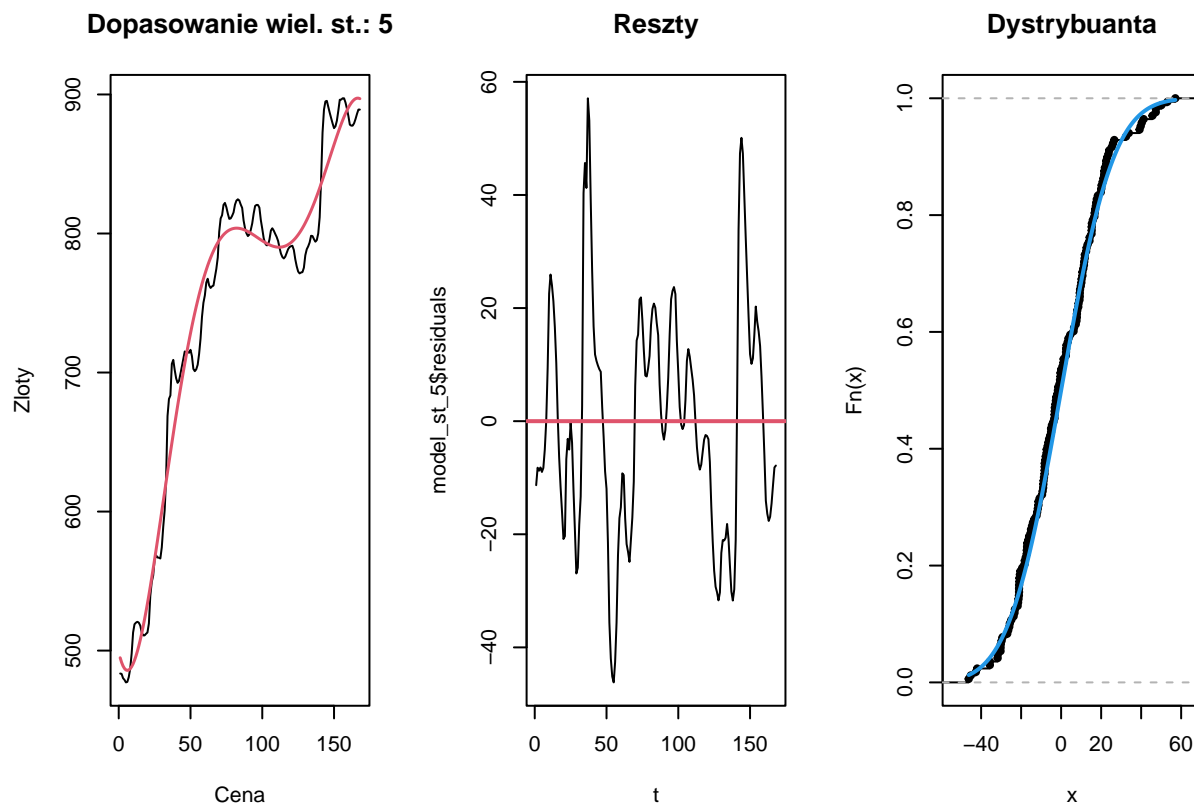


Z kryterium osuwiska wybieramy wielomian stopnia 5.

```
par(mfrow = c(1, 3))
model_st_5 <- lm(dane ~ ., data=ramka)
plot(t, dane, type = "l",
     main = paste("Dopasowanie wiel. st.:", 5),
     ylab = "Złoty", xlab = "Cena")
lines(t, model_st_5$fitted.values, col = 2, lwd = 1.5)

plot(t, model_st_5$residuals, main = "Reszty", type = "l")
abline(h = 0, col = 2, lwd = 2)

plot(ecdf(model_st_5$residuals), main = "Dystrybuanta")
x <- seq(from = min(model_st_5$residuals), to = max(model_st_5$residuals), length.out = 500)
lines(x, pnorm(x, mean = 0, sd = sd(model_st_5$residuals)), col = 4, lwd = 2)
```

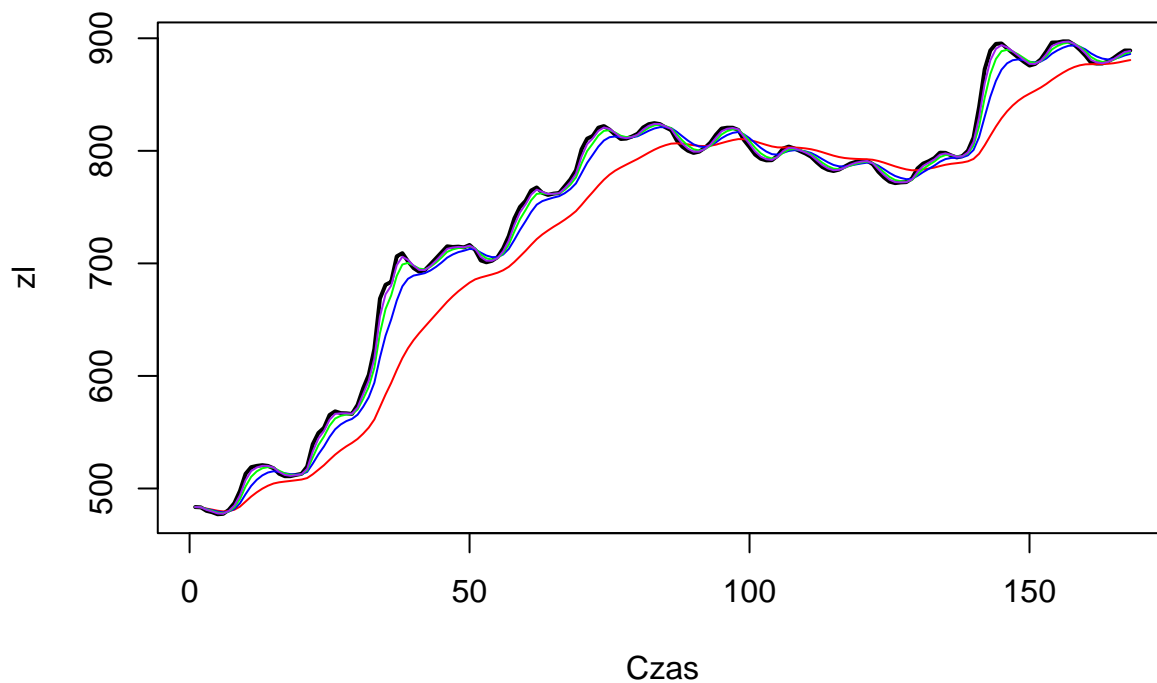


```
par(mfrow = c(1, 1))
```

Metoda wykładniczych wag ruchomej średniej

```
plot(dane, main = "Ceny węgla kamiennego", lwd = 2, type = "l", xlab = "Czas", ylab = "zł")
wykładnicza(dane, 0.9, "red")
wykładnicza(dane, 0.7, "blue")
wykładnicza(dane, 0.5, "green")
wykładnicza(dane, 0.3, "purple")
```

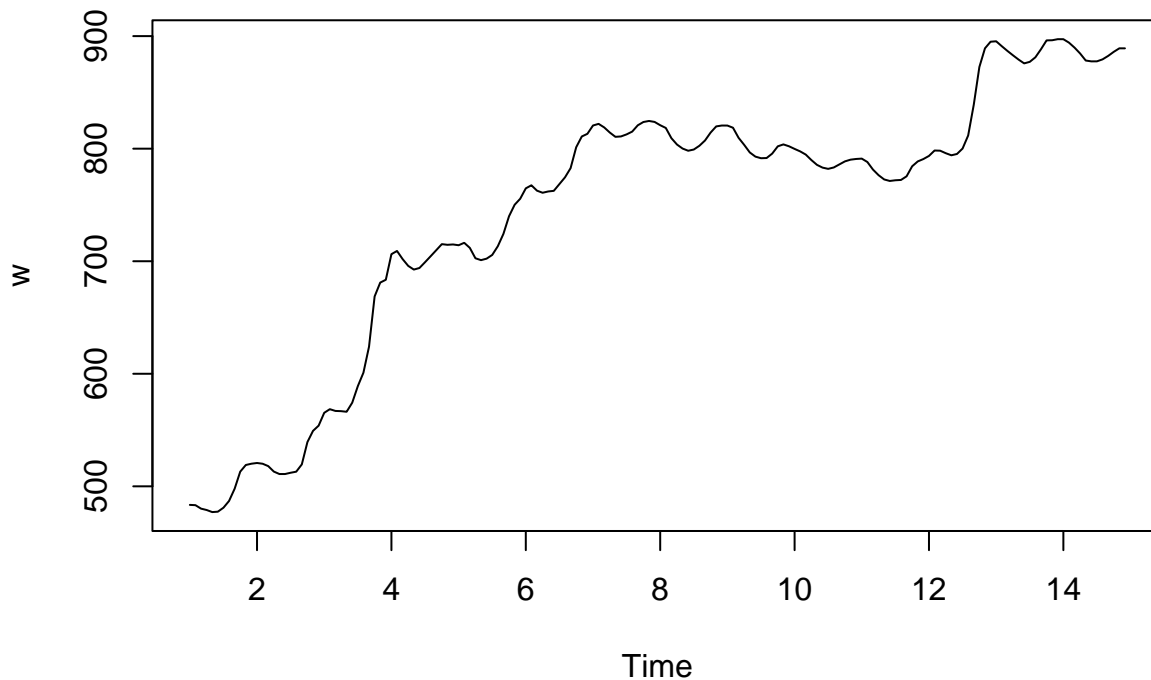
## Ceny węgla kamiennego



W celu eliminacji losowych fluktuacji w szeregu czasowym użyliśmy metody wykładniczych wag ruchomej średniej. Widać, że wraz ze wzrostem współczynnika wag funkcja przybliżająca mniej naśladuje zaburzenia zewnętrzne w modelu i na odwrót.

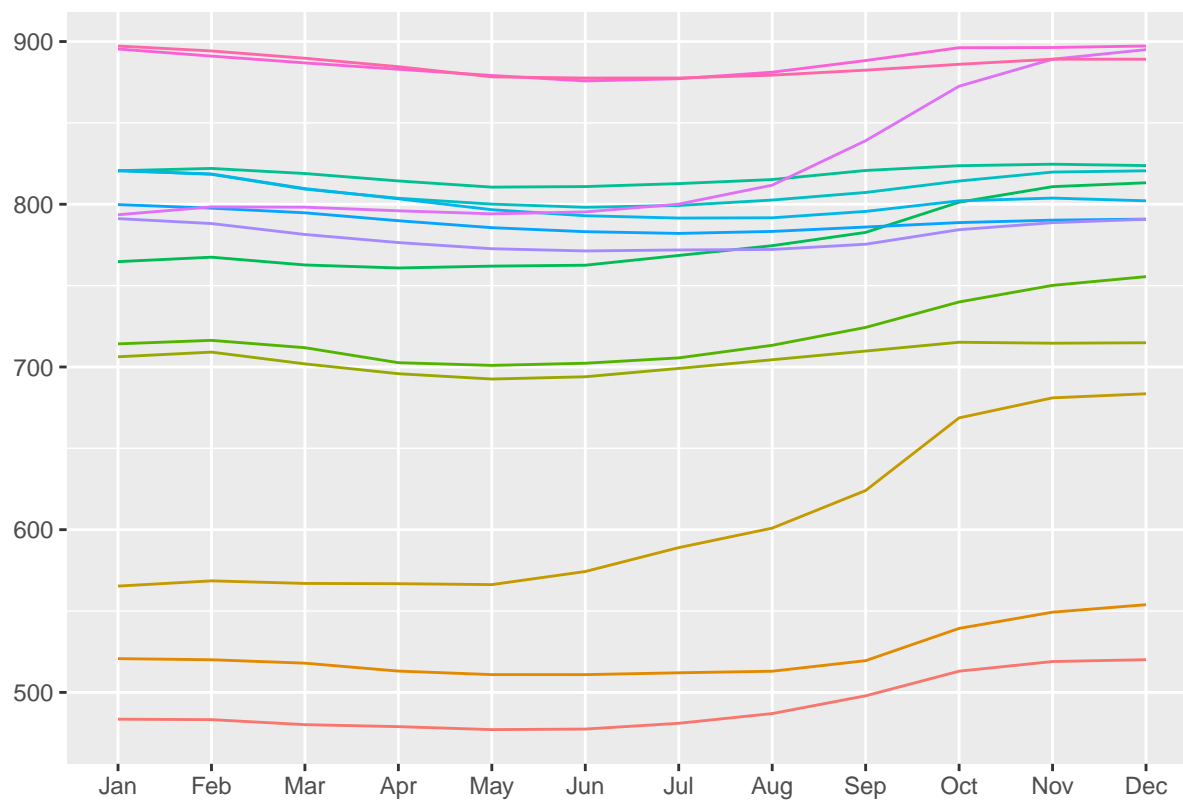
### Analiza trendów fazowych

```
w <- ts(dane, frequency = 12)
plot(w, type = "l")
```



```
ggseasonplot(w, polar = FALSE, main = "Wykres sezonowości dla lat 2006-2019", xlab = NULL) +  
  theme(legend.position = "none")
```

Wykres sezonowosci dla lat 2006-2019



Z wykresu sezonowości widzimy powtarzający się trend wzrostu cen węgla kamiennego w okresie od sierpnia do listopada. W okresie od stycznia do maja zauważalny jest nieznaczny trend spadkowy cen.