

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("wagiel_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)  
typeof(macierz[1,2])
```

```
## [1] "double"
```

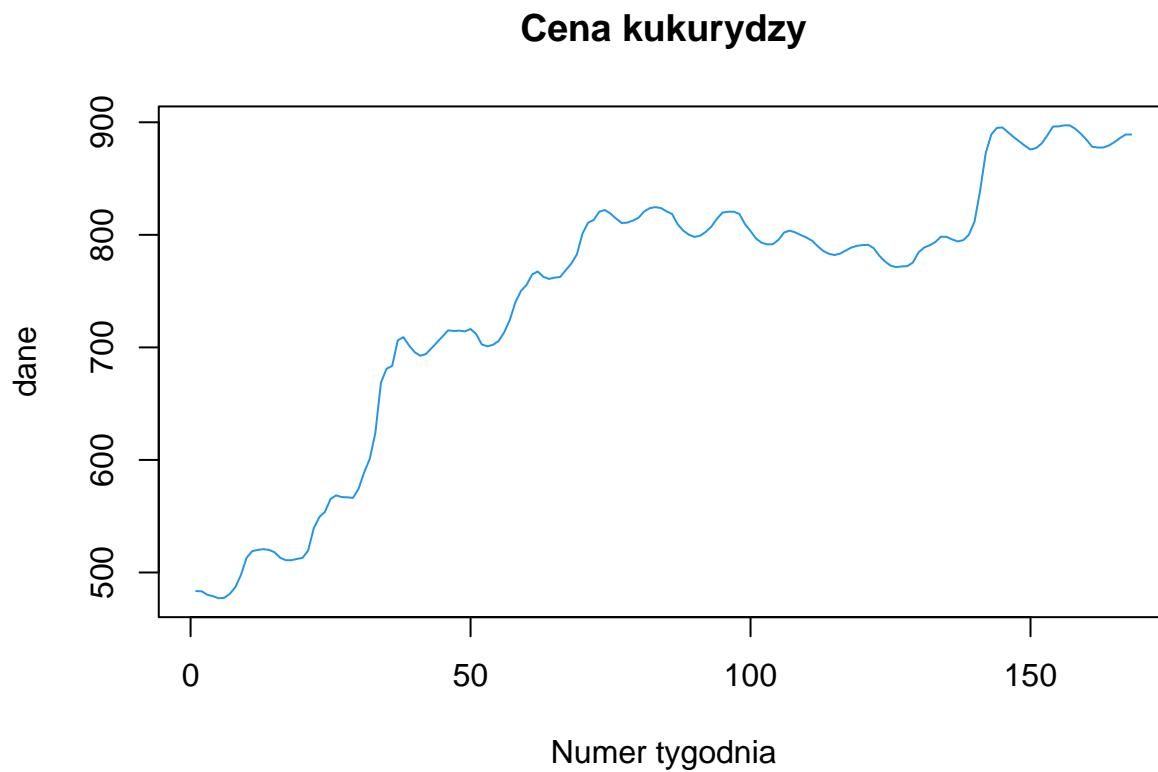
Przekształcenie macierzy w wektor czytany kolumnowo (od góry do dołu). W ten sposób otrzymujemy dane w odpowiedniej kolejności.

```
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.

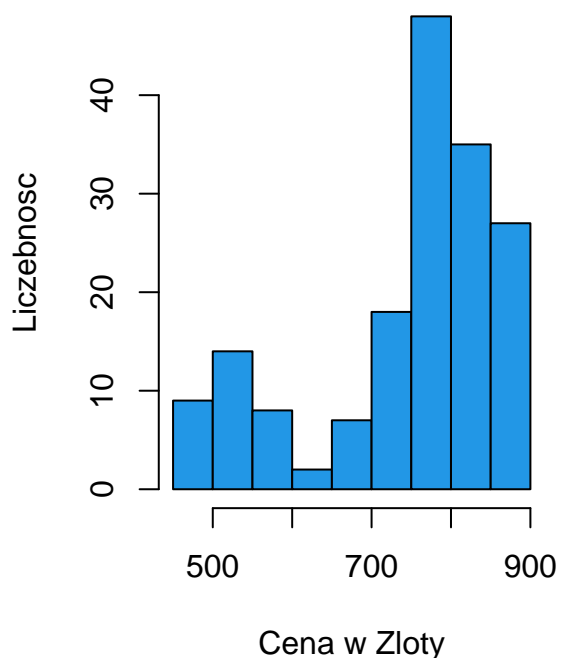
Robimy wykresy typu boxplot oraz histogram, żeby zobaczyć rozkład danych. Wykresy typu boxplot oraz histogram.

```
par(mfrow = c(1, 2))

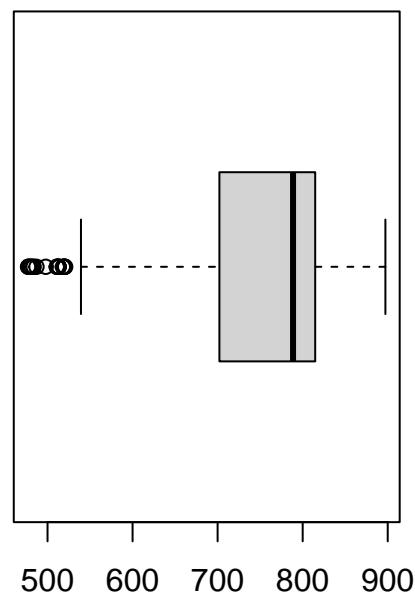
hist(dane, breaks = 15, col = 4, main = "Histogram cen węgla kamiennego",
     xlab = "Cena w Złoty", ylab = "Liczebność", prob = FALSE)

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 kwartyl pierwszy wynosi około 700 a kwartyl 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
```

```
Akaike <- c()
```

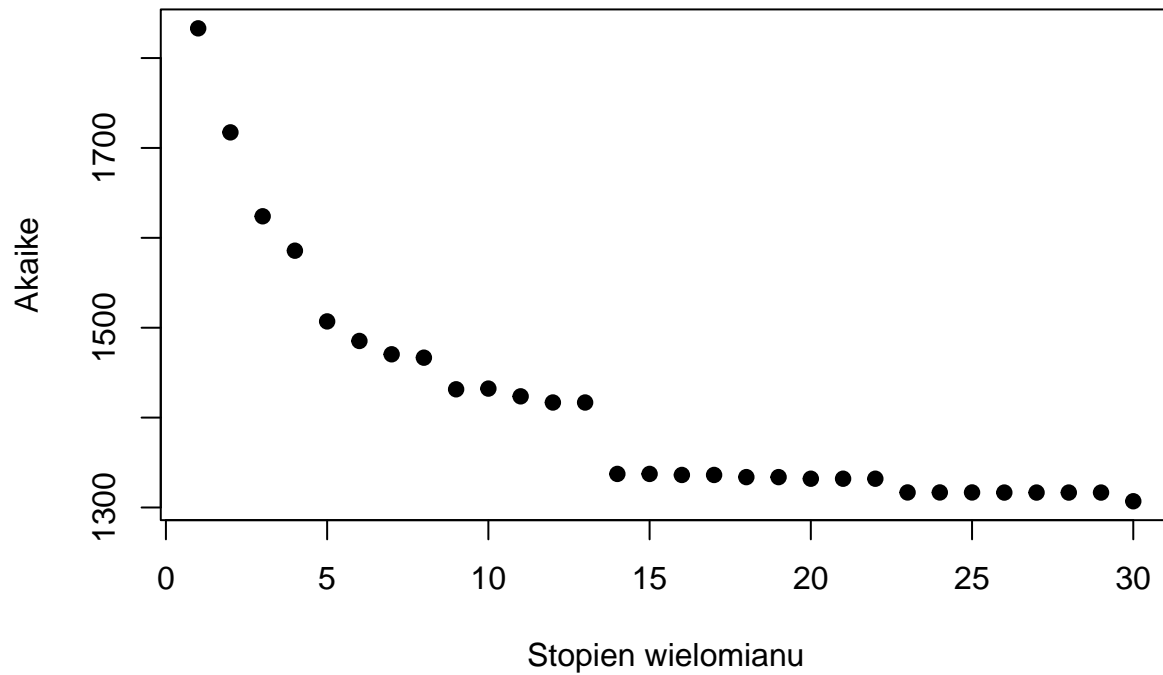
```
for(i in 1:30){
```

```
  Akaike <- cbind(Akaike, dopasowanie_wielomianu(dane,i))
```

```
}
```

```
plot(1:30, Akaike,type="p", pch=19, main = "Kryterium AIC dla wielomianu stopnia i", xlab = "Stopień wielomianu")
```

Kryterium AIC dla wielomianu stopnia i



```
Akaike[c(10:15)]
```

```
## [1] 1432.278 1423.643 1416.808 1416.808 1337.352 1337.352
```

```
par(mfrow = c(1, 3)) plot(t, szereg, type = "l", main = paste("Dopasowanie wiel. st.:", stopien), ylab = "Złoty", xlab = "Cena") lines(t, model$fitted.values, col = 2, lwd = 1.5)
```

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

```
plot(ecdf(model$residuals), main = "Dystrybuanta") x <- seq(from = min(model$residuals), to = max(model$residuals), length.out = 500) lines(x, pnorm(x, mean = 0, sd = sd(model$residuals)), col = 4, lwd = 2) par(mfrow = c(1, 1))
```