

SPRAWOZDANIE

Zajęcia: Nauka o danych I

Prowadzący: prof. dr hab. Vasyl Martsenyuk

Laboratorium Nr 9 Data 25.01.2025 Temat: "9. Wprowadzenie do analizy czasowych szeregów danych 10. Projektowanie eksperymentów i testowanie hipotez w praktyce" Wariant 7	Tomasz Pietrzyk Informatyka II stopień, niestacjonarne, 1semestr, gr.1a
---	--

1. Polecenie: wariant 7 zadania

7. Przeanalizuj dane o miesięcznej produkcji energii elektrycznej w USA:

- Pobierz dane z [Time Series Datasets](#).
- Oblicz średnie ruchome dla okien 6 i 12 miesięcy.
- Przeprowadź analizę ACF i PACF.
- Zdekomponuj dane na składniki trendu, sezonowości i reszt.

2. Opis programu opracowanego (kody źródłowe, rzuty ekranu)

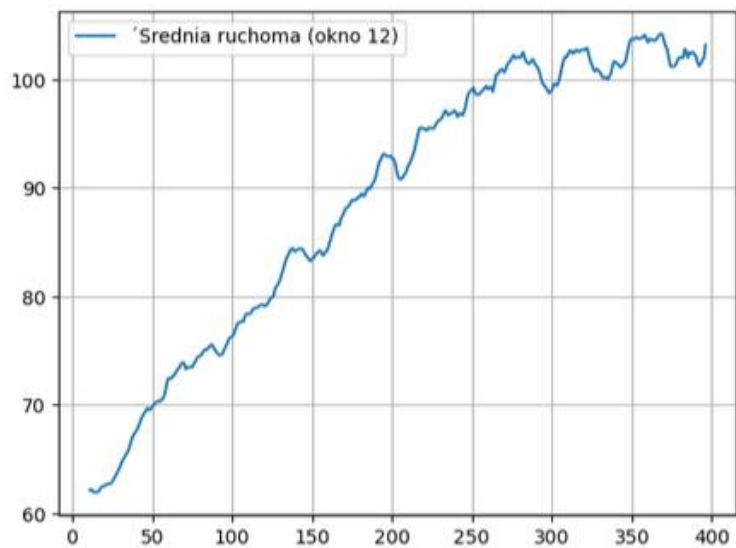
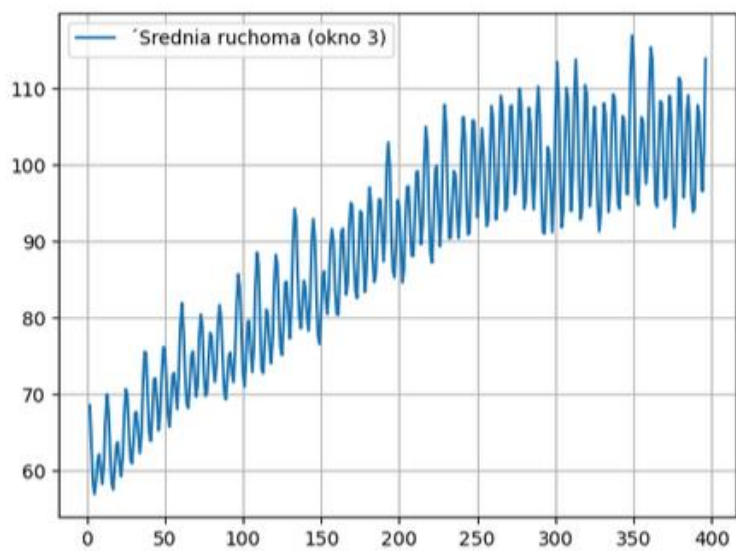
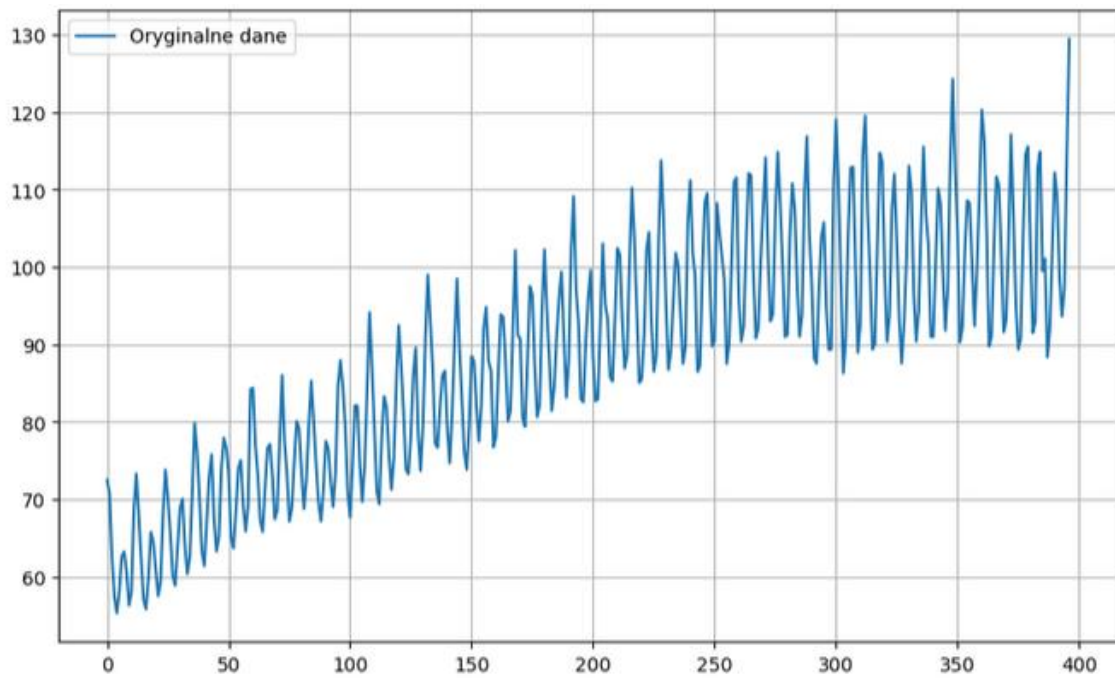
GitHub: [https://github.com/TomekPietrzyk/NOD I 2024 NS.git](https://github.com/TomekPietrzyk/NOD_I_2024_NS.git)

```
[47]: import pandas as pd
import matplotlib.pyplot as plt

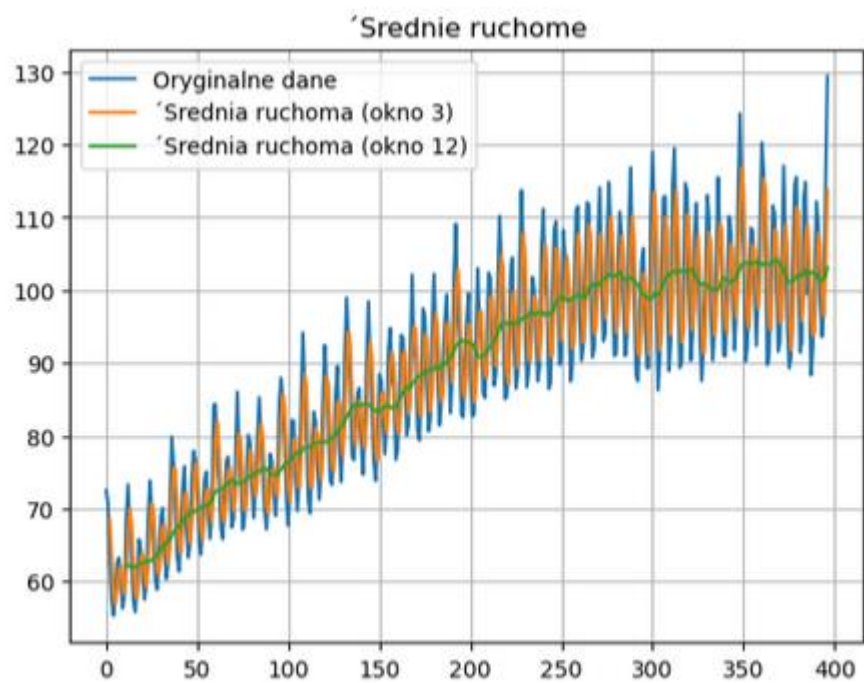
df = pd.read_csv("Electric_Production.csv")
print(df.head())
df.info()

# Obliczanie 'srednich ruchomych'
df['MA_3'] = df['IPG2211A2N'].rolling(window=3).mean()
df['MA_12'] = df['IPG2211A2N'].rolling(window=12).mean()
df

# Wizualizacja
plt.figure(figsize=(10, 6))
plt.plot(df['IPG2211A2N'], label="Oryginalne dane")
plt.legend()
plt.grid()
plt.show()
plt.plot(df['MA_3'], label="Srednia ruchoma (okno 3)")
plt.legend()
plt.grid()
plt.show()
plt.plot(df['MA_12'], label="Srednia ruchoma (okno 12)")
plt.legend()
plt.grid()
plt.show()
plt.plot(df['IPG2211A2N'], label="Oryginalne dane")
plt.plot(df['MA_3'], label="Srednia ruchoma (okno 3)")
plt.plot(df['MA_12'], label="Srednia ruchoma (okno 12)")
plt.title("Srednie ruchome")
plt.legend()
plt.grid()
plt.show()
```



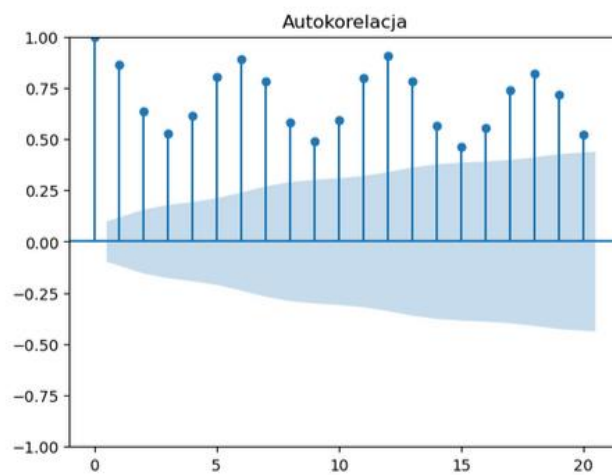
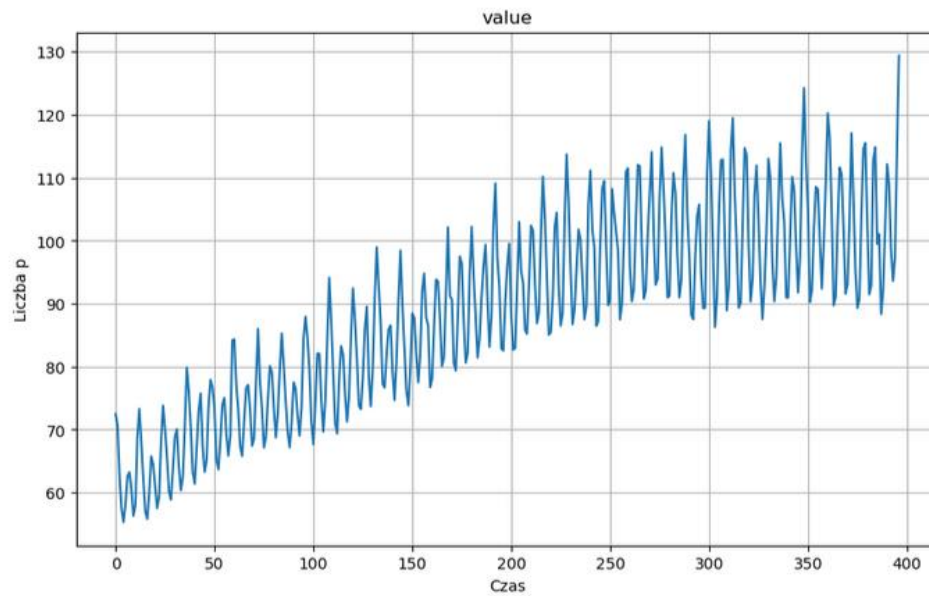
'Srednia ruchoma

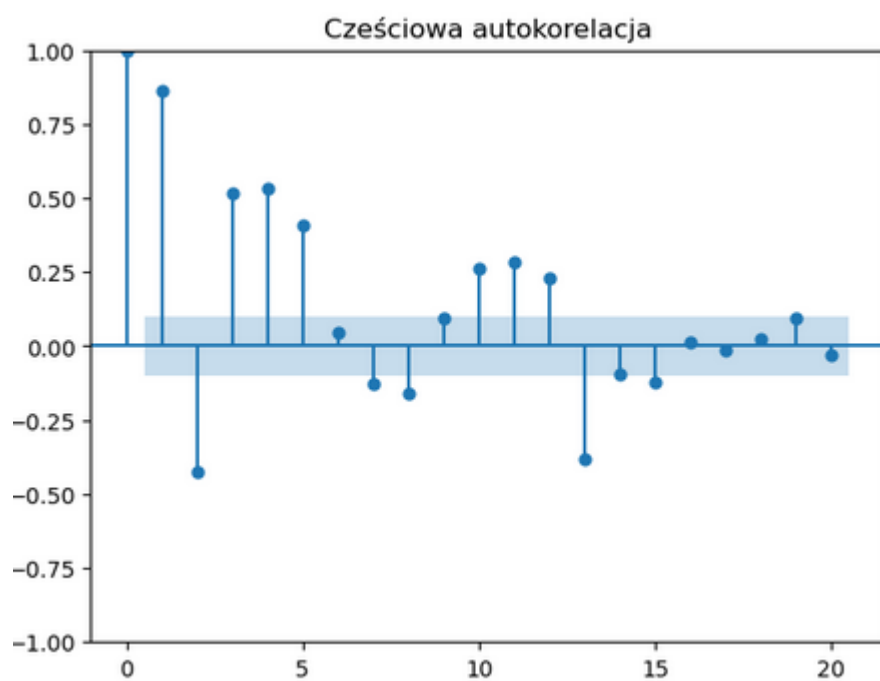


```
[61]: #ACF
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf

df = df.rename(columns={'IPG2211A2N' : 'value'})
plt.figure(figsize=(10, 6))
df['value'].plot(title="value", xlabel="Czas", ylabel="Liczba p")
plt.grid()
plt.show()

plot_acf(df['value'], lags=20, title="Autokorelacja")
plot_pacf(df['value'], lags=20, title="Cze'sciowa autokorelacja")
plt.show()
```





```
[125]: import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.seasonal import seasonal_decompose
df = pd.read_csv("Electric_Production.csv")
df = df.rename(columns={"IPG221A2N": "value"})

# Konwersja kolumny 'DATE' na format datetime
df['DATE'] = pd.to_datetime(df['DATE'])

# Ustawienie indeksu na kolumnę 'DATE'
df.set_index('DATE', inplace=True)

# Sprawdzenie poprawności indeksu
print(df.info())

# Ustawienie częstotliwości (jeśli dane są dzienne)
# df = df.asfreq('H', method='pad') # Uzupełnia brakujące wartości poprzednimi danymi

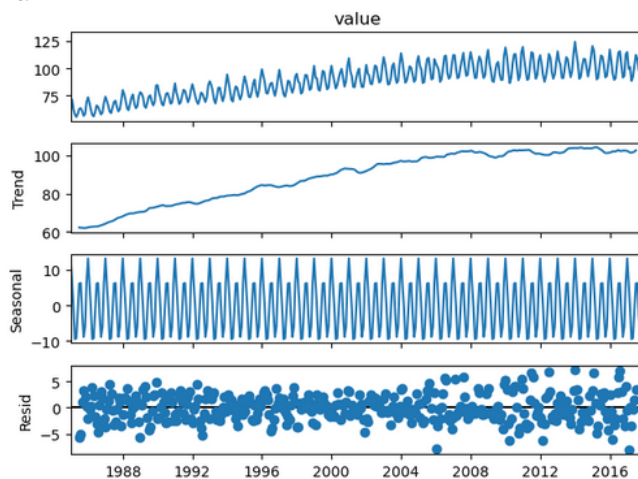
print(df.info())

# Sprawdzenie, czy są braki w danych
print(df.isna().sum())

# Dekonpozycja - WAŻNE: określamy period!
result = seasonal_decompose(df['value'], model='additive') # Okres 7 dla danych dziennych

# Wizualizacja
result.plot()
plt.show()
```

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 397 entries, 1985-01-01 to 2018-01-01
Data columns (total 1 columns):
#   Column  Non-Null Count  Dtype
---  --
0   value    397 non-null        float64
dtypes: float64(1)
memory usage: 6.2 KB
None
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 397 entries, 1985-01-01 to 2018-01-01
Data columns (total 1 columns):
#   Column  Non-Null Count  Dtype
---  --
0   value    397 non-null        float64
dtypes: float64(1)
memory usage: 6.2 KB
None
value      0
dtype: int64
```



3. Wnioski

Na podstawie zebranych danych można sugerować, że trend jest rosnący oraz zachodzi swego rodzaju sezonowość. Z analizy wykresów ACF i PACF można stwierdzić, że szereg czasowy wykazuje silną autokorelację w początkowych lagach, co wskazuje na obecność trendu lub niestacjonarności w danych. Wartości ACF powoli zanikają, co sugeruje potrzebę różnicowania danych w celu

usunięcia trendu. Z kolei PACF pokazuje istotną korelację dla pierwszego i ewentualnie drugiego lagu, co może wskazywać na proces autoregresyjny rzędu 1 lub 2.